Erratum: Quenched large deviation principle for words in a letter sequence

M. Birkner [1]
A. Greven [2]
F. den Hollander [3]

October 1, 2022

Abstract

In [BGdH10] we derived a quenched large deviation principle for the empirical process of words obtained by cutting an i.i.d. sequence of letters according to an independent renewal process. We derived a representation of the associated rate function for stationary word processes in terms of certain specific relative entropies. Our proof of this representation is correct when the mean word length is finite, but is flawed when the mean word length is infinite. In this paper we fix the flaw in the proof. Along the way we derive new representations of the rate function that are interesting in their own right. A key ingredient in the proof is the observation that if the rate function in the annealed large deviation principle is finite at a stationary word process, then the letters in the tail of the long words in this process are typical.

MSC2010: 60F10, 60G10.
Keywords: Letters and words, Renewal process, Empirical process, Annealed vs. quenched, Large deviation principle, Rate function, Specific relative entropy.
Acknowledgment: We are grateful to Jean-Christophe Mourrat for pointing out the flaw in the original proof. We thank Evgeny Verbitskiy for fruitful discussions on ergodic theory.

1 Large deviation principles

Sections 1.1–1.2 (which are largely copied from [BGdH10, Sections 1.1–1.2]) state the original theorem. Section 1.3 explains at what spot the original proof is flawed and how the flaw can be fixed. Section 1.4 provides further perspectives, lists the papers where the original theorem was used, and outlines the remainder of the paper.

1.1 Problem setting

Let E be a Polish space. View the elements of E as *letters*. Let $\widetilde{E} = \cup_{n\in\mathbb{N}} E^n$ be the set of finite *words* drawn from E, which is also a Polish space (see e.g. Remark A.9 in Appendix A; in

[1] Institut für Mathematik, Johannes-Gutenberg-Universität, Staudingerweg 9, 55099 Mainz, Germany, *birkner@mathematik.uni-mainz.de*
[2] Department Mathematik, Universität Erlangen-Nürnberg, Cauerstrasse 11, D-91058 Erlangen, Germany, *greven@mi.uni-erlangen.de*
[3] Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands, *denholla@math.leidenuniv.nl*

case $|E| < \infty$, we can simply use the discrete topology). Let $\mathcal{P}(E^{\mathbb{N}})$ and $\mathcal{P}(\widetilde{E}^{\mathbb{N}})$ denote the set of probability measures on sequences drawn from E, respectively, $\widetilde{E}$, equipped with the topology of weak convergence. Write θ and $\widetilde{\theta}$ for the left-shift acting on $E^{\mathbb{N}}$, respectively, $\widetilde{E}^{\mathbb{N}}$. Write $\mathcal{P}^{\mathrm{inv}}(E^{\mathbb{N}}), \mathcal{P}^{\mathrm{erg}}(E^{\mathbb{N}})$ and $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}), \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})$ for the set of probability measures that are invariant, respectively, invariant and ergodic under θ, respectively, $\widetilde{\theta}$.

For $\nu \in \mathcal{P}(E)$, let $X = (X_i)_{i\in\mathbb{N}}$ be i.i.d. with marginal law ν. Without loss of generality we will assume that $\mathrm{supp}(\nu) = E$ (otherwise we replace E by $\mathrm{supp}(\nu)$). For $\rho \in \mathcal{P}(\mathbb{N})$, let $\tau = (\tau_i)_{i\in\mathbb{N}}$ be i.i.d. with marginal law ρ having infinite support and satisfying the *algebraic tail property*

$$\lim_{\substack{n\to\infty \\ \rho(n)>0}} \frac{\log \rho(n)}{\log n} =: -\alpha, \quad \alpha \in (1,\infty). \tag{1.1}$$

(The parameter α plays an important role in the sequel. No regularity assumption is necessary for $\mathrm{supp}(\rho)$.) Assume that X and τ are independent and write $\mathbb{P}$ to denote their joint law. Cut words out of X according to τ, i.e., put (see Fig. 1)

$$T_0 := 0, \qquad T_i := T_{i-1} + \tau_i, \quad i \in \mathbb{N}, \tag{1.2}$$

and let

$$Y^{(i)} := \big(X_{T_{i-1}+1}, X_{T_{i-1}+2}, \dots, X_{T_i}\big), \quad i \in \mathbb{N}. \tag{1.3}$$

Then, under the law $\mathbb{P}$, $Y = (Y^{(i)})_{i\in\mathbb{N}}$ is an i.i.d. sequence of words with marginal law $q_{\rho,\nu}$ on $\widetilde{E}$ given by

$$\begin{aligned} &q_{\rho,\nu}\big((x_1,\dots,x_n)\big) := \mathbb{P}\big(Y^{(1)} = (x_1,\dots,x_n)\big) = \rho(n)\,\nu(x_1)\cdots\nu(x_n), \\ &n \in \mathbb{N},\ x_1,\dots,x_n \in E. \end{aligned} \tag{1.4}$$

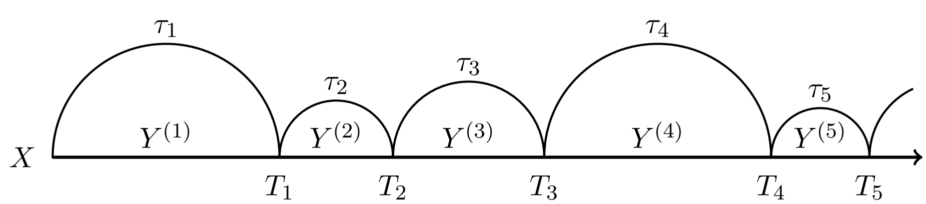

Figure 1: Cutting words from a letter sequence according to a renewal process.

For $N \in \mathbb{N}$, let $(Y^{(1)},\dots,Y^{(N)})^{\mathrm{per}}$ stand for the periodic extension of $(Y^{(1)},\dots,Y^{(N)})$ to an element of $\widetilde{E}^{\mathbb{N}}$, and define

$$R_N := \frac{1}{N} \sum_{i=0}^{N-1} \delta_{\widetilde{\theta}^i (Y^{(1)},\dots,Y^{(N)})^{\mathrm{per}}} \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}), \tag{1.5}$$

called the *empirical process of N-tuples of words*. By the ergodic theorem, we have

$$\mathrm{w}-\lim_{N\to\infty} R_N = q_{\rho,\nu}^{\otimes\mathbb{N}} \quad \mathbb{P}\text{–a.s.}, \tag{1.6}$$

with $\mathrm{w}-\lim$ denoting the weak limit.

1.2 Quenched LDP

To state the quenched LDP for $\mathbb{P}(R_N \in \cdot \mid X)$, $N \in \mathbb{N}$, we require some more notation. Define

$$\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}}) = \{Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})\colon\ m_Q < \infty\}, \tag{1.7}$$

where

$$m_Q := \mathbb{E}_Q[\tau_1] \tag{1.8}$$

is the mean word length of Q. Let $\kappa\colon \widetilde{E}^{\mathbb{N}} \to E^{\mathbb{N}}$ denote the *concatenation map* that glues a sequence of words into a sequence of letters. For $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$, define $\Psi_Q \in \mathcal{P}^{\mathrm{inv}}(E^{\mathbb{N}})$ as

$$\Psi_Q(\cdot) := \frac{1}{m_Q}\mathbb{E}_Q\left[\sum_{k=0}^{\tau_1-1} \delta_{\theta^k\kappa(Y)}(\cdot)\right]. \tag{1.9}$$

Think of Ψ_Q as *the shift-invariant version of the concatenation of Y under the law Q obtained after randomising the location of the origin.* In fact, when $m_Q < \infty$, $\kappa(Q) := Q \circ \kappa^{-1} \in \mathcal{P}(E^{\mathbb{N}})$ is asymptotically mean stationary with stationary mean Ψ_Q. (See [CdH13, Lemma 5.1]. Also see [G11] for general properties of asymptotically mean stationary measures. In the notation of [G11], (1.9) reads $\Psi_Q = \overline{\kappa(Q)}$.)

For $\mathrm{tr} \in \mathbb{N}$, let $[\cdot]_{\mathrm{tr}}\colon\ \widetilde{E} \to [\widetilde{E}]_{\mathrm{tr}} := \cup_{n=1}^{\mathrm{tr}} E^n$ denote the *word length truncation* map defined by

$$y = (x_1, \dots, x_n) \mapsto [y]_{\mathrm{tr}} := (x_1, \dots, x_{n\wedge \mathrm{tr}}), \qquad n \in \mathbb{N},\ x_1, \dots, x_n \in E. \tag{1.10}$$

Extend this to a map from $\widetilde{E}^{\mathbb{N}}$ to $[\widetilde{E}]_{\mathrm{tr}}^{\mathbb{N}}$ via

$$\big[(y^{(1)}, y^{(2)}, \dots)\big]_{\mathrm{tr}} := \big([y^{(1)}]_{\mathrm{tr}}, [y^{(2)}]_{\mathrm{tr}}, \dots\big) \tag{1.11}$$

and to a map from $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ to $\mathcal{P}^{\mathrm{inv}}([\widetilde{E}]_{\mathrm{tr}}^{\mathbb{N}})$ via

$$[Q]_{\mathrm{tr}}(A) := Q(\{z \in \widetilde{E}^{\mathbb{N}}\colon\ [z]_{\mathrm{tr}} \in A\}), \qquad A \subset [\widetilde{E}]_{\mathrm{tr}}^{\mathbb{N}} \text{ measurable.} \tag{1.12}$$

Note that if $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$, then $[Q]_{\mathrm{tr}} \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ for all $\mathrm{tr} \in \mathbb{N}$.

Write $H(\cdot \mid \cdot)$ to denote *specific relative entropy* on $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ or $\mathcal{P}^{\mathrm{inv}}(E^{\mathbb{N}})$, i.e., per word or per letter, respectively.

Theorem 1.1. [Quenched LDP: original version, [BGdH10, Theorem 1.2 and Corollary 1.6]] *Assume* (1.1). *Then, for $\nu^{\otimes\mathbb{N}}$–a.s. all X, the family of (regular) conditional probability distributions $\mathbb{P}(R_N \in \cdot \mid X)$, $N \in \mathbb{N}$, satisfies the LDP on $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ with rate N and with deterministic rate function $I^{\mathrm{que}}\colon\ \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) \to [0,\infty]$ given by*

$$I^{\mathrm{que}}(Q) := \begin{cases} I^{\mathrm{fin}}(Q), & \text{if } Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}}),\\ \lim\limits_{\mathrm{tr}\to\infty} I^{\mathrm{fin}}\big([Q]_{\mathrm{tr}}\big), & \text{otherwise,} \end{cases} \tag{1.13}$$

where

$$I^{\mathrm{fin}}(Q) := H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) + (\alpha-1)\, m_Q\, H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}). \tag{1.14}$$

Moreover, I^{que} has compact level sets, has a unique zero at $Q = q_{\rho,\nu}^{\otimes\mathbb{N}}$, is affine, and is equal to the lower semi-continuous extension of I^{fin} from $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ to $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$.

The proof in [BGdH10] of the representation of $I^{\rm que}$ in the second line of (1.13) contains a *flaw*, as pointed out to the authors by Jean-Christophe Mourrat (private communication; see also [M17, p. 25–26]). The goal of the present paper is to fix this flaw. It will become clear from what follows that fixing the proof is a delicate matter. The proofs in [BGdH10] of the other claims in Theorem 1.1 are correct.

Remark 1.2. [Support of ρ] In view of the generality allowed in (1.1), strictly speaking in the second line of (1.13) we take $\mathrm{tr} \to \infty$ only along subsequences where $\rho(\mathrm{tr}) > 0$. Otherwise, $I^{\rm fin}\big([Q]_{\rm tr}\big)$ would be ∞ for the trivial reason that under $[Q]_{\rm tr}$ words of length tr appear, which would be impossible under $q_{\rho,\nu}^{\otimes\mathbb{N}}$ for tr's with $\rho(\mathrm{tr}) = 0$. We silently use the convention that $\lim_{\mathrm{tr}\to\infty}$ means $\lim_{\mathrm{tr}\to\infty,\,\rho(\mathrm{tr})>0}$, and analogously for $\limsup$ and $\liminf$. $\triangle$

1.3 The missing piece

Let $d_{\mathcal{P}^{\rm inv}(\widetilde{E}^{\mathbb{N}})}$ be a suitable metric on $\mathcal{P}^{\rm inv}(\widetilde{E}^{\mathbb{N}})$ that metrises weak convergence. Write

$$B_\varepsilon(Q) := \big\{Q' \in \mathcal{P}^{\rm inv}(\widetilde{E}^{\mathbb{N}})\colon d_{\mathcal{P}^{\rm inv}(\widetilde{E}^{\mathbb{N}})}(Q,Q') < \varepsilon\big\} \tag{1.15}$$

for the open ε-ball around $Q \in \mathcal{P}^{\rm inv}(\widetilde{E}^{\mathbb{N}})$. [BGdH10, Propositions 3.1 and 4.1] show that, for every $Q \in \mathcal{P}^{\rm inv,fin}(\widetilde{E}^{\mathbb{N}})$ with $H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) < \infty$ and every $\delta > 0$, there is an $\varepsilon = \varepsilon(Q,\delta) > 0$ such that X-a.s.

$$\begin{aligned} -I^{\rm fin}(Q) - \delta &\le \liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big(R_N \in R_N \in B_\varepsilon(Q) \mid X\big) \\ &\le \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big(R_N \in R_N \in \overline{B_\varepsilon(Q)} \mid X\big) \le -I^{\rm fin}(Q) + \delta. \end{aligned} \tag{1.16}$$

Thus, in [BGdH10] the following version of Theorem 1.1 was proved but not explicitly stated.

Theorem 1.3. [Quenched LDP: proved version from [BGdH10]] *Assume* (1.1). *Then, for $\nu^{\otimes\mathbb{N}}$–a.s. all X, the family of (regular) conditional probability distributions $\mathbb{P}(R_N \in \cdot \mid X)$, $N \in \mathbb{N}$, satisfies the LDP with rate N and with deterministic rate function $\widetilde{I}^{\rm que}\colon \mathcal{P}^{\rm inv}(\widetilde{E}^{\mathbb{N}}) \to [0,\infty]$ given by*

$$\widetilde{I}^{\rm que}(Q) = \sup_{\varepsilon>0} \inf\big\{I^{\rm fin}(Q')\colon\, Q' \in \mathcal{P}^{\rm inv,fin}(\widetilde{E}^{\mathbb{N}}) \cap B_\varepsilon(Q)\big\}, \quad Q \in \mathcal{P}^{\rm inv}(\widetilde{E}^{\mathbb{N}}). \tag{1.17}$$

Moreover, $\widetilde{I}^{\rm que}$ has compact level sets, has a unique zero at $Q = q_{\rho,\nu}^{\otimes\mathbb{N}}$, is convex, equals $I^{\rm fin}$ on $\mathcal{P}^{\rm inv,fin}(\widetilde{E}^{\mathbb{N}})$, is affine on $\mathcal{P}^{\rm inv,fin}(\widetilde{E}^{\mathbb{N}})$, and is equal to the lower semi-continuous extension of $I^{\rm fin}$ from $\mathcal{P}^{\rm inv,fin}(\widetilde{E}^{\mathbb{N}})$ to $\mathcal{P}^{\rm inv}(\widetilde{E}^{\mathbb{N}})$.

Proof. For finite alphabets, the proof that the family $\mathbb{P}(R_N \in \cdot \mid X)$, $N \in \mathbb{N}$, satisfies the LDP with rate function $\widetilde{I}^{\rm que}$ given by (1.17), and that $\widetilde{I}^{\rm que}$ has compact level sets, follows from [BGdH10, Section 5.1, Steps 1a–1c] by using (1.16) and the fact that $\mathcal{P}^{\rm inv,fin}(\widetilde{E}^{\mathbb{N}})$ is a dense subset of $\mathcal{P}^{\rm inv}(\widetilde{E}^{\mathbb{N}})$. As shown in [BGdH10, Section 6, Step 2], $I^{\rm fin}$ is lower semi-continuous on $\mathcal{P}^{\rm inv,fin}(\widetilde{E}^{\mathbb{N}})$, and so (1.17) shows that $\widetilde{I}^{\rm que} = I^{\rm fin}$ on $\mathcal{P}^{\rm inv,fin}(\widetilde{E}^{\mathbb{N}})$, and that $\widetilde{I}^{\rm que}$ is the lower semi-continuous extension of $I^{\rm fin}$ from $\mathcal{P}^{\rm inv,fin}(\widetilde{E}^{\mathbb{N}})$ to $\mathcal{P}^{\rm inv}(\widetilde{E}^{\mathbb{N}})$. The extension to Polish alphabets was given in [BGdH10, Section 8], and the arguments given there for the case of finite mean word length are correct. We revisit the case of infinite mean word length for general Polish alphabets in Section 2.7 below.

Affineness of I^{fin} on $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ was proved in [BGdH10, Section 6, Step 1], and together with (1.17) this implies convexity of $\widetilde{I}^{\mathrm{que}}$: For $Q = \lambda Q_1 + (1-\lambda)Q_2$ with $Q_1, Q_2 \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$, $\lambda \in (0,1)$, let $(Q_{i,n})_{n\in\mathbb{N}} \subset \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ be such that $\mathrm{w}-\lim_{n\to\infty} Q_{i,n} = Q_i$ and $\lim_{n\to\infty} I^{\mathrm{fin}}(Q_{i,n}) = \widetilde{I}^{\mathrm{que}}(Q_i)$ for $i = 1,2$. Then $Q = \mathrm{w}-\lim_{n\to\infty}(\lambda Q_{1,n} + (1-\lambda)Q_{2,n})$, and hence $\widetilde{I}^{\mathrm{que}}(Q) \leq \liminf_{n\to\infty}(\lambda I^{\mathrm{fin}}(Q_{1,n}) + (1-\lambda)I^{\mathrm{fin}}(Q_{2,n})) = \lambda \widetilde{I}^{\mathrm{que}}(Q_1) + (1-\lambda)\widetilde{I}^{\mathrm{que}}(Q_1)$. □

Thus, we are left with fixing the representation of the rate function given in the second line of (1.13) and verifying that I^{que} is affine on all of $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$. For this we need to turn to the *truncated empirical process*. Fix a truncation level $\mathrm{tr} \in \mathbb{N}$. In [BGdH10, Section 5, p. 442] it is *erroneously* claimed that the family $[R_N]_{\mathrm{tr}}$, $N \in \mathbb{N}$, of truncated empirical processes satisfies the LDP with rate function $I^{\mathrm{fin}}([Q]_{\mathrm{tr}})$ (see Remark 1.6 below). The correct statement for fixed tr is the following.

Lemma 1.4. [Quenched LDP: truncated version] *Let* $\mathrm{tr} \in \mathbb{N}$. *The family* $\mathbb{P}([R_N]_{\mathrm{tr}} \in \cdot \mid X)$, $N \in \mathbb{N}$, *satisfies the LDP on*

$$\mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}}) = \big\{Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) \colon\ Q(|Y^{(1)}| \leq \mathrm{tr}) = 1\big\} \subset \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) \tag{1.18}$$

with rate N and with deterministic rate function $\widetilde{I}_{\mathrm{tr}} \colon \mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}}) \to [0,\infty]$ *given by*

$$\widetilde{I}_{\mathrm{tr}}(Q) \ = \inf\big\{\widetilde{I}^{\mathrm{que}}(Q') :\ Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}), [Q']_{\mathrm{tr}} = Q\big\}, \qquad Q \in \mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}}). \tag{1.19}$$

An alternative representation of the rate function reads

$$\widetilde{I}_{\mathrm{tr}}(Q) = \inf\big\{I^{\mathrm{fin}}(Q') :\ Q' \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}}), [Q']_{\mathrm{tr}} = Q\big\}. \tag{1.20}$$

Proof. Theorem 1.3 and the contraction principle (see e.g. [DZ98, Theorem 4.2.1]) imply that $\mathbb{P}([R_N]_{\mathrm{tr}} \in \cdot \mid X)$, $N \in \mathbb{N}$, satisfies the LDP with rate function $\widetilde{I}_{\mathrm{tr}}(Q)$ given by (1.19). The alternative representation in (1.20) is proved in Appendix C.1. The proof makes use of (2.12), which requires Theorem 2.3 below. We will not use (1.20) in this paper, in particular, it is not needed for the proof of Theorem 2.3. □

From Lemma 1.4 and the argument in [BGdH10, Section 5, Step 2] we see that

$$\widetilde{I}^{\mathrm{que}}(Q) = \sup_{\mathrm{tr}\in\mathbb{N}} \widetilde{I}_{\mathrm{tr}}([Q]_{\mathrm{tr}}) = \lim_{\mathrm{tr}\to\infty} \widetilde{I}_{\mathrm{tr}}([Q]_{\mathrm{tr}}), \qquad Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}). \tag{1.21}$$

Indeed, the first equality follows from Lemma 1.4 and the Dawson-Gärtner projective limit LDP, which also guarantees that $\widetilde{I}^{\mathrm{que}}$ has the properties of a rate function (see e.g. [DZ98, Theorem 4.6.1]), while the second equality follows from the fact that $\widetilde{I}_{\mathrm{tr}+1}([Q]_{\mathrm{tr}+1}) \geq \widetilde{I}_{\mathrm{tr}}([Q]_{\mathrm{tr}})$ (observe that any Q' with $[Q']_{\mathrm{tr}+1} = [Q]_{\mathrm{tr}+1}$ automatically satisfies $[Q']_{\mathrm{tr}} = [Q]_{\mathrm{tr}}$). From the above observations, we have that

$$\widetilde{I}^{\mathrm{que}}(Q) \leq \liminf_{\mathrm{tr}\to\infty} I^{\mathrm{fin}}([Q]_{\mathrm{tr}}), \quad Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}), \tag{1.22}$$

because $\widetilde{I}^{\mathrm{que}}$ is lower semi-continuous and $\widetilde{I}_{\mathrm{tr}}([Q]_{\mathrm{tr}}) \leq I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) = \widetilde{I}^{\mathrm{que}}([Q]_{\mathrm{tr}})$. Using [BGdH10, Lemma A.1], we also know that $\widetilde{I}^{\mathrm{que}}(Q) = \lim_{\mathrm{tr}\to\infty} I^{\mathrm{fin}}([Q]_{\mathrm{tr}})$ for $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$. However, at this point of the discussion we are not yet able to rule out the possibility that there is a sequence $(Q'_n)_{n\in\mathbb{N}} \subset \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ with $Q'_n \to^w Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) \setminus \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ (with $\to^w$ denoting weak convergence) for which

$$\lim_{n\to\infty} I^{\mathrm{fin}}(Q'_n) < \liminf_{\mathrm{tr}\to\infty} I^{\mathrm{fin}}([Q]_{\mathrm{tr}}). \tag{1.23}$$

In view of the above discussion, in order to *salvage* the representation of the rate function in the second line of (1.13), we need that

$$\limsup_{\mathrm{tr}\to\infty} I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) \le \widetilde{I}^{\mathrm{que}}(Q), \qquad Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}). \tag{1.24}$$

Note that (1.24) does hold (with equality) on $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$, and is plausible on $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})\backslash\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ because, by the definition of weak convergence, in order for R_N to be close to Q it must get the statistics of all the short words approximately right. It is plausible that the most parsimonious way to achieve this approximation is to arrange that R_N is close to $[Q]_{\mathrm{tr}}$ for some large tr. In fact, this intuition is corroborated in a quantitative way by Lemma 2.13 below (see also Remark A.8 for a more elementary version): For Q with $I^{\mathrm{ann}}(Q) < \infty$, the letter content of "long" words (longer than tr) is close to being $\nu^{\otimes\mathbb{N}}$-typical and hence in the limit as $\mathrm{tr}\to\infty$ does not contribute to the quenched rate function.

Our main result in this paper is the following.

Theorem 1.5. **[Flaw fixed]** (1.24) *is true. Furthermore,* I^{que} *is affine on all of* $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$*.*

We will first prove Theorem 1.5 for finite alphabets E and ergodic measures Q, and afterwards pass to Polish alphabets via a projective limit argument and to non-ergodic measures via an ergodic decomposition argument.

1.4 Discussion

Remark 1.6. **[Properties of the truncated rate function]** Note that, unlike I^{fin}, $\widetilde{I}_{\mathrm{tr}}$ from Lemma 1.4 satisfies $\widetilde{I}_{\mathrm{tr}} \equiv \infty$ on $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) \setminus \mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$, because $[R_N]_{\mathrm{tr}}$ takes values in $\mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$ a.s. To see why it is too much to hope for an explicit formula like

$$H(Q \mid [q_{\rho,\nu}^{\otimes\mathbb{N}}]_{\mathrm{tr}}) + (\alpha-1) m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) \tag{1.25}$$

for $Q \in \mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$, as is *erroneously* claimed in [BdH13, Corollary 2.2], consider $E = \{0,1\}$, $\nu = (1-p)\delta_0 + p\delta_1$, $\mathrm{tr} = 1$ and $Q = (\delta_{(1)})^{\otimes\mathbb{N}}$. Then $\Psi_Q = (\delta_1)^{\otimes\mathbb{N}}$, and $\mathbb{P}([R_N]_1 \approx Q)$ amounts to the question studied in [BGdH10, Lemma 2.1] in the asymptotic regime $p \downarrow 0$. Equation (1.25) would give the explicit value $\alpha\log(1/p)$ for the exponential decay rate of the event from [BGdH10, Lemma 2.1]. While this is the correct asymptotics in the limit as $p\downarrow 0$, it is wrong for fixed p. In particular, it does not depend on any detail of ρ other than the tail parameter α. See Section C.2 for more details. $\triangle$

Remark 1.7. **[Annealed LDP]** The annealed LDP stated in [BGdH10, Theorem 1.1] is standard (see e.g. Dembo and Zeitouni [DZ98, Corollaries 6.5.15 and 6.5.17]). The associated rate function is $I^{\mathrm{ann}}(Q) = H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}})$, the *specific relative entropy of* Q *w.r.t.* $q_{\rho,\nu}^{\otimes\mathbb{N}}$. This rate function is lower semi-continuous, has compact level sets, has a unique zero at $Q = q_{\rho,\nu}^{\otimes\mathbb{N}}$, and is affine. In [BGdH10, (1.17)] it is correctly claimed that $I^{\mathrm{ann}}(Q) = \lim_{\mathrm{tr}\to\infty} I^{\mathrm{ann}}([Q]_{\mathrm{tr}})$ for $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$, i.e., in the annealed LDP there is no problem with the truncation limit. $\triangle$

Remark 1.8. **[Problems]** The following statements in [BGdH10] were dubious:

(1) In the proof of [BGdH10, Theorem 1.2] for finite alphabets given in [BGdH10, Section 5], in Step 2, Eq. (5.8) is literally wrong and in Step 3, Eqs. (5.10)–(5.11) do not follow from the arguments given there.

(2) In the proof of [BGdH10, Corollary 1.6] in Section 8, the argument given for Eq. (8.15) is not correct.

(3) In the proof of [BGdH10, Theorem 1.3], Step 4 in Section 6, the argument as it stands does in general not allow to transfer affineness from I^{fin} to I^{que}.

In these instances, we replaced a supremum over $\mathrm{tr} \in \mathbb{N}$ by a limit $\mathrm{tr} \to \infty$ without proper justification.

(4) Furthermore, in [BGdH10, Section 5, p. 442] below Eq. (5.1), as discussed in Lemma 1.4 and Remark 1.6, we made an incorrect claim about the rate function of the truncated empirical process.

In the present paper, we rectify these problems. $\triangle$

Remark 1.9. [Applications] The quenched LDP with i.i.d. letters stated in Theorem 1.1 was used in [BGdH11], [BdH13], [CdH13], [dHO13], [BdHO15], [N18], [CdH21]. The first six explicitly mention the formulation of the quenched rate function via the truncation approximation, while the seventh uses only the lower semi-continuous regularization from (1.17). In [BGdH11], [dHO13], [BdHO15] and [N18], only (1.13) is mentioned or used and therefore no further amendment is necessary. In [CdH13, Eq. (2.12)], (1.22) is erroneously used with I^{que} instead of $\widetilde{I}^{\mathrm{que}}$. The incorrect claim that $I^{\mathrm{que}}(Q) = \sup_{\mathrm{tr}} I^{\mathrm{que}}([Q]_{\mathrm{tr}})$ is used only in the proof of [CdH13, Lemma 3.6]. For $m_Q < \infty$ the proof is complete. For $m_Q = \infty$ part (i) is proved in this erratum (see in particular Sections 2.7–2.8), the proofs of parts (ii) and (iii) can be adapted analogously. The arxiv preprint [BdH13] also states the incorrect claim but in fact only uses the limit formula from (1.13).

A quenched LDP with a general weak Bernoulli letter sequence appears in [dHP14]. The proof of LDP in [dHP14, Theorems 1.2–1.3] is correct, but the limit formula [dHP14, Eq. (1.8)] will need an extension of the arguments in the present paper from i.i.d. letters to weak Bernoulli letters. This extension is plausible but non-trivial. $\triangle$

Outline. The proof of Theorem 1.5 is given in Section 2, which consists of 8 subsections forming the bulk of the paper. This proof is based on an abstract manipulation of entropies and relative entropies, in combination with a number of technical estimates. In Appendix A we collect a few auxiliary facts and observations that are needed in Section 2, while in Appendix B we discuss and adapt some ergodic theoretic constructions that are used in Section 2, Appendix C provides proofs of additional remarks and observations.

2 Proof of Theorem 1.5

Section 2.1 collects a number of preliminary observations that are needed along the way. Section 2.2 contains an outline of the proof of Theorem 1.5 for ergodic measures and finite alphabets, which consists of 5 Steps. Section 2.3 works out these steps in detail, based on a number of technical lemmas. Sections 2.4–2.6 use these lemmas to complete the proof. Section 2.7 extends the proof to Polish alphabets, Section 2.8 to non-ergodic measures.

2.1 Preliminary observations

Remarks 2.1–2.2 below contain several preliminary observations that are needed along the way and that will help the reader to appreciate the subtleties behind the proof of Theorem 1.5.

Remark 2.1. [Comparison of rate functions] Fix $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ with $m_Q = \infty$. Probabilistically, (1.24) says that it is not harder under $\mathbb{P}(\cdot \mid X)$ for R_N to be $\approx [Q]_{\mathrm{tr}}$ than to be in any small neighbourhood of Q when $\mathrm{tr} \gg 1$. Note that, from Theorem 1.3 and general facts about large deviation principles, we have

$$\begin{aligned} -I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) &= \lim_{\epsilon \downarrow 0} \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in B_\epsilon([Q]_{\mathrm{tr}}) \mid X) \\ &= \lim_{\epsilon \downarrow 0} \liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in B_\epsilon([Q]_{\mathrm{tr}}) \mid X), \\ -\widetilde{I}^{\mathrm{que}}(Q) &= \lim_{\epsilon \downarrow 0} \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in B_\epsilon(Q) \mid X) \\ &= \lim_{\epsilon \downarrow 0} \liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in B_\epsilon(Q) \mid X), \end{aligned} \tag{2.1}$$

and so (1.24) is *equivalent* to the condition that for any $\delta > 0$ there is a $\mathrm{tr}_0 = \mathrm{tr}_0(Q,\delta)$ such that, for all $\mathrm{tr} \geq \mathrm{tr}_0$,

$$\begin{aligned} &\lim_{\epsilon \downarrow 0} \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in B_\epsilon([Q]_{\mathrm{tr}}) \mid X) \\ &\qquad \geq -\delta + \lim_{\epsilon \downarrow 0} \liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in B_\epsilon(Q) \mid X). \end{aligned} \tag{2.2}$$

This in turn is equivalent to the condition that, for any $\delta > 0$ there are $\mathrm{tr}_0 = \mathrm{tr}_0(Q,\delta)$ and sequences $(\varepsilon_k)_{k\in\mathbb{N}}$, $(\varepsilon'_k)_{k\in\mathbb{N}}$ tending to zero such that

$$\liminf_{N\to\infty} e^{\delta N} \frac{\mathbb{P}(R_N \in B_{\varepsilon_k}([Q]_{\mathrm{tr}}) \mid X)}{\mathbb{P}(R_N \in B_{\varepsilon'_k}(Q) \mid X)} \geq 1 \quad \forall\, \mathrm{tr} \geq \mathrm{tr}_0(Q,\delta),\ k \geq k_0 = k_0(Q,\mathrm{tr}). \tag{2.3}$$

Note the joint quantification of k and tr in (2.3): For fixed $\varepsilon > 0$, we can arrange that $B_\varepsilon(Q) \subset B_{2\varepsilon}([Q]_{\mathrm{tr}})$ for all $\mathrm{tr} \geq \mathrm{tr}_0(Q,\varepsilon)$. It is necessary to be allowed to let $\varepsilon \downarrow 0$ uniformly in tr for sufficiently large tr.

The possible non-uniformity is part of the problem. If we would know that for every $\delta > 0$ and $\varepsilon_0 > 0$ we can find $\varepsilon \in (0, \varepsilon_0]$ and $\mathrm{tr}_0 = \mathrm{tr}_0(Q,\delta,\varepsilon)$ such that

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in B_\varepsilon([Q]_{\mathrm{tr}}) \mid X) \leq -I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) + \delta \quad \forall\, \mathrm{tr} \geq \mathrm{tr}_0, \tag{2.4}$$

then we would obtain (1.24). We could even allow $\varepsilon = \varepsilon(\mathrm{tr}) \downarrow 0$ as long as $d(Q, [Q]_{\mathrm{tr}}) < \varepsilon(\mathrm{tr})$. Indeed, assume (2.4). Then for every $\varepsilon > 0$ we can choose tr so large that $Q \in B_\varepsilon([Q]_{\mathrm{tr}})$, and so

$$-\widetilde{I}^{\mathrm{que}}(Q) \leq \liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in B_\varepsilon([Q]_{\mathrm{tr}}) \mid X) \leq -I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) + \delta \tag{2.5}$$

for all tr sufficiently large, and hence $\widetilde{I}^{\mathrm{que}}(Q) \geq \limsup_{\mathrm{tr}\to\infty} I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) - \delta$. (We will show below that all these conditions are in fact fulfilled.) $\triangle$

Remark 2.2. [Alternative representations for $I^{\rm fin}$] For $Q \in \mathcal{P}^{\rm inv,fin}(\widetilde{E}^{\mathbb{N}})$,

$$\begin{aligned} I^{\rm que}(Q) &= H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) + (\alpha-1) m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) \\ &= \alpha H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) + (\alpha-1)\Big(\mathbb{E}_Q\big[\log(\rho(\tau_1))\big] + H_{\tau|K}(Q)\Big) \\ &= \alpha m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) - \mathbb{E}_Q\big[\log(\rho(\tau_1))\big] - H_{\tau|K}(Q), \end{aligned} \tag{2.6}$$

where we combine (1.14) with the relation

$$H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) = m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) - \mathbb{E}_Q\big[\log(\rho(\tau_1))\big] - H_{\tau|K}(Q) \tag{2.7}$$

from [B08, Lemma 4] (see [BGdH10, Eq. (1.32)]). Now consider $Q \in \mathcal{P}^{\rm inv}(\widetilde{E}^{\mathbb{N}})$ with $m_Q = \infty$ and $I^{\rm ann}(Q) = H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) < \infty$. Since $H([Q]_{\rm tr} \mid q_{\rho,\nu}^{\otimes\mathbb{N}}) \to H(Q \mid q_{\rho,\nu}^{\otimes\mathbb{N}})$ as ${\rm tr} \to \infty$ (recall Remark 1.7) and

$$\mathbb{E}_{[Q]_{\rm tr}}\big[\log(\rho(\tau_1))\big] = \mathbb{E}_Q\big[\log(\rho(\tau_1 \wedge {\rm tr}))\big] \underset{{\rm tr}\to\infty}{\longrightarrow} \mathbb{E}_Q\big[\log(\rho(\tau_1))\big] \in (-\infty, 0) \tag{2.8}$$

by dominated convergence (note that $|\log\rho({\rm tr})| \le c \log {\rm tr}$ and $\mathbb{E}_Q[\log(\tau_1)] < \infty$ by Lemma A.2), we see that

$$\begin{aligned} &\lim_{{\rm tr}\to\infty} m_{[Q]_{\rm tr}} H(\Psi_{[Q]_{\rm tr}} \mid \nu^{\otimes\mathbb{N}}) \text{ exists} \iff \lim_{{\rm tr}\to\infty} I^{\rm que}([Q]_{\rm tr}) \text{ exists} \\ &\iff \lim_{{\rm tr}\to\infty} H_{\tau|K}([Q]_{\rm tr}) \text{ exists} \end{aligned} \tag{2.9}$$

(use (2.7) for the second equivalence). We will see below that these limits do indeed exist. △

2.2 Structure of the proof

The key to the proof of Theorem 1.5 is the following theorem.

Theorem 2.3. [Truncation limit is asymptotically optimal] *Let $Q \in \mathcal{P}^{\rm inv}(\widetilde{E}^{\mathbb{N}})$ with $I^{\rm ann}(Q) < \infty$. For every $\delta > 0$ there exists a ${\rm tr}_0 = {\rm tr}_0(Q,\delta)$ such that for every truncation level ${\rm tr} \ge {\rm tr}_0$ there is an open set $\mathcal{O} = \mathcal{O}(Q,{\rm tr}) \subset \mathcal{P}^{\rm inv}(\widetilde{E}^{\mathbb{N}})$ with $Q \in \mathcal{O}$ (and also $[Q]_{{\rm tr}'} \in \mathcal{O}$ for every ${\rm tr}' \ge {\rm tr}$) such that*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in \mathcal{O} \mid X) \le -I^{\rm fin}([Q]_{\rm tr}) + \delta \quad a.s. \tag{2.10}$$

Theorem 2.3 implies Theorem 1.5. Indeed,

$$-\widetilde{I}^{\rm que}(Q) \le \liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in \mathcal{O} \mid X) \quad \text{a.s.} \tag{2.11}$$

by Theorem 1.3, which together with (2.10) implies (1.24). Note that Theorem 2.3 also implies that

$$\lim_{{\rm tr}\to\infty} I^{\rm fin}([Q]_{\rm tr}) = I^{\rm que}(Q) \tag{2.12}$$

exists (and of course equals $\widetilde{I}^{\rm que}(Q)$ from (1.17) in Theorem 1.3).

Having established the limit formula (1.13), we see that $I^{\rm que}$ is affine because, for every ${\rm tr} \in \mathbb{N}$, the mapping $Q \mapsto I^{\rm fin}([Q]_{\rm tr})$ is affine (cf. Theorem 1.3).

In the remainder of this section we explain how we prove Theorem 2.3. The general idea behind the proof is the following. The condition $I^{\rm ann}(Q) < \infty$ enforces that the content of *long words* (of length $\geq$ tr) must be almost $\nu^{\otimes\mathbb{N}}$-typical (see Lemma 2.13 below, and also Remark A.8, for a quantified version). Consequently, on the exponential scale the price for producing the long words in the given realisation of the letter sequence X is almost negligible.

The proof hinges on two aims, namely, express

(1) $H_{\tau|K}([Q]_{\mathrm{tr}})$

(2) $m_{[Q]_{\mathrm{tr}}} H(\Psi_{[Q]_{\mathrm{tr}}} \mid \nu^{\otimes\mathbb{N}})$

in terms of characteristics of $[Q]_{\mathrm{tr}}$ that involve only the *short words* (of length $<$ tr) up to a controlled error. (Note that these aims are related: using (2.33) from Lemma 2.10 together with the relation (2.7) from Remark 2.2, each aim can in principle be transformed into the other. See also the heuristics on p. 410 of [BGdH10] explaining how these two quantities appear in the large deviation probability.) Equipped with Aims (1) and (2), the proof of the upper bound for the case $m_Q < \infty$ given in [BGdH10, Sections 3.2–3.4] can be adapted, as we explain in Sections 2.3–2.6 below.

The proof of Theorem 2.3 partly mimics and partly extends the proof of the upper bound for the case $m_Q < \infty$ given in [BGdH10]. In fact, it provides an alternative way to prove [BGdH10, Proposition 3.1]. A crucial difference is that we express the constituent terms of $I^{\mathrm{fin}}([Q]_{\mathrm{tr}})$, namely, the terms in Aims (1) and (2) above (cf. (2.6), in Remark 2.2) in terms of characteristics of $[Q]_{\mathrm{tr}}$ that involve only the short words up to a controlled error (see Proposition 2.15 below), which achieves Aims (1) and (2). By contrast, when working directly with words drawn from $[Q]_{\mathrm{tr}}$ as we did in [BGdH10], long words and their letter content have in principle to be taken into account. Note that even though long words are rare, when $m_Q = \infty$ they typically contribute a non-trivial fraction of the letters under $\Psi_{[Q]_{\mathrm{tr}}}$ for any tr $< \infty$.

We start the proof by assuming that $|E| < \infty$ and $Q \in \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})$. Here is an outline of the five main steps that are needed to achieve Aims (1) and (2), which are detailed in Section 2.3.

- Section 2.3.1 defines excursions of short words separated by long words, which leads to an induced process of short words that plays a crucial role in the construction.
- Section 2.3.2 looks at the concatenation $\widehat{\kappa}_{\mathrm{tr}}$ of words inside each excursion. The crucial point is that the letter content of the resulting process $\widehat{\kappa}([Q]_{\mathrm{tr}})$ is the same as that of $[Q]_{\mathrm{tr}}$ (see Lemma 2.7), while the "unwieldy" quantity $H_{\tau|K}([Q]_{\mathrm{tr}})$ is replaced by something that is asymptotically negligible as tr $\to \infty$. In general, computing the specific entropy of a concatenated word process is hard and the relation to the specific entropy of the original word process is highly non-explicit (compare with the problem of computing the specific entropy of a hidden Markov chain). However, the discrepancy becomes less important the longer the words are (see (2.22) in Lemma 2.8 for a quantitative version).
- In Section 2.3.3 we replace long words by a special symbol $*_{\geq \mathrm{tr}}$. This provides an alternative way of approximating both Q and the rate function at Q in terms of $\varphi_{\mathrm{tr}}(Q)$ rather than $[Q]_{\mathrm{tr}}$ (see Lemma 2.10). Via Corollary 2.11, this is also the key tool for the proof of Lemma 2.13, which suitably quantifies the idea that long words are almost $\nu^{\otimes \mathrm{tr}}$-typical.
- Section 2.3.4 shows that the same operations for $\varphi_{\mathrm{tr}}(Q)$ (excursion decomposition and concatenation of words inside the excursions) yield objects that depend only on the statistics of words of length $<$ tr under Q. Proposition 2.15 expresses both $m_{[Q]_{\mathrm{tr}}} H(\Psi_{[Q]_{\mathrm{tr}}} \mid \nu^{\otimes\mathbb{N}})$ and $H_{\tau|K}([Q]_{\mathrm{tr}})$ in terms of these objects, up to controlled error.

- Section 2.3.5 turns the various estimates obtained into quantitative error estimates for specific relative entropies that are needed in Sections 2.4–2.6 to complete the proof of Theorem 1.5 for finite alphabets and ergodic measures.

Armed with these results, we complete the proof of Theorem 2.3 as follows. In Section 2.4 we construct a specific neighbourhood of Q and of $[Q]_{\mathrm{tr}}$ for large tr that is well adapted to formalise the strategy outlined above. In Section 2.5, we define the empirical process of words obtained by cutting the letter sequence at given locations, and identify the cuttings for which the empirical process falls into this neighbourhood. In Section 2.6, we estimate the cost of finding suitable stretches in the letter sequence for which this mapping can be done properly.

In Section 2.7 we extend the proof to Polish E, and in Section 2.8 to non-ergodic $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$.

Remark 2.4. [Alternative representation of the rate function] It will become clear from the proof of Theorem 2.3 that, in addition to the representation (1.13), we also have (at least in the case of at most countable E)

$$I^{\mathrm{que}}(Q) = \lim_{\mathrm{tr}\to\infty} \Bigg(H\big(\varphi_{\mathrm{tr}}(Q) \,\big|\, \varphi_{\mathrm{tr}}(q_{\rho,\nu}^{\otimes\mathbb{N}})\big) - (\alpha-1)\varepsilon_{\mathrm{tr}} H\big(\widehat{\kappa}_{\mathrm{tr}}^{*}(Q)\big) \\ - (\alpha-1)\mathbb{E}_Q\Big[\mathbb{1}_{\{|Y^{(1)}|<\mathrm{tr}\}} \sum_{j=1}^{|Y^{(1)}|} \log \nu(Y_j^{(1)})\Big]\Bigg). \tag{2.13}$$

See (2.31), Lemma 2.10 and (2.40), Lemma 2.12 below for the definitions of the objects appearing in (2.13). See also Proposition 2.15. Note that the terms inside the large brackets on the right-hand side (2.13) depend only on the statistics under Q of words with length $<$ tr.

As it stands, (2.13) literally only works for $|E| < \infty$. However, the two terms $-\varepsilon_{\mathrm{tr}} H(\widehat{\kappa}_{\mathrm{tr}}^{*}(Q)) - \mathbb{E}_Q[\mathbb{1}_{\{|Y^{(1)}|<\mathrm{tr}\}} \sum_{j=1}^{|Y^{(1)}|} \log \nu(Y_j^{(1)})]$ taken together have an interpretation as the specific relative entropy of the letter content of the short words with respect to $\nu^{\otimes\mathbb{N}}$, which is meaningful for general E. $\triangle$

2.3 Objects and operations

We abbreviate $Q_{\rho,\nu} := q_{\rho,\nu}^{\otimes\mathbb{N}}$. Throughout Sections 2.3–2.6 we assume that $|E| < \infty$, $Q \in \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})$ with $m_Q = \infty$, and $H(Q \mid Q_{\rho,\nu}) < \infty$. Sections 2.3.1–2.3.5 carry out the five main steps mentioned above.

2.3.1 Excursion decomposition and induced process

The idea is to decompose a (finite or infinite) sequence $y = (y_i)_{i\in\mathbb{N}}$ of words into (finite) excursions (= stretches) according to the appearance of long words that have $\mathrm{tr} \in \mathbb{N}$ or more letters. To formalize the decomposition, define a mapping $\widehat{\varrho}_{\geq\mathrm{tr}} : \widetilde{E}^{\mathbb{N}} \to (\widehat{E}^{(\mathrm{tr})})^{\mathbb{N}}$, where $\widehat{E}^{(\mathrm{tr})} = E^{\geq\mathrm{tr}} \times \cup_{m\in\mathbb{N}_0}(E^{<\mathrm{tr}})^m$ with $E^{\geq\mathrm{tr}} = \cup_{\ell\geq\mathrm{tr}} E^{\ell}$, $E^{<\mathrm{tr}} = \cup_{\ell=1}^{\mathrm{tr}-1} E^{\ell}$. For $y = (y^{(n)})_{n\in\mathbb{N}} \in \widetilde{E}^{\mathbb{N}}$, write $t_i = t_i(y) = \inf\{n > t_{i-1} \colon |y^{(n)}| \geq \mathrm{tr}\}$, $i \in \mathbb{N}$, with $t_0 := 0$, and put

$$\widehat{y}^{(i)} := \big(y^{(t_i)}, y^{(t_i+1)}, \dots, y^{(t_{i+1}-1)}\big), \qquad i \in \mathbb{N}. \tag{2.14}$$

Here, we implicitly assume that y has the property that $|y^{(i)}| \geq \mathrm{tr}$ for infinitely many i's. We think of this operation as cutting the path $(y^{(n)})_{n\in\mathbb{N}}$ into excursions away from the set $E^{\geq\mathrm{tr}}$. By our notation convention, each such excursion begins with a long word (from $E^{\geq \mathrm{tr}}$) and is followed by zero or more short words (from $E^{<\mathrm{tr}}$). For $w = (w^{(1)}, \dots, w^{(m+1)}) \in \widehat{E}^{(\mathrm{tr})}$, $m \in \mathbb{N}_0$, we write $\mathsf{len}(w) = m+1$ for the length of w (in words).

Note that we can recover $y = (y^{(n)})_{n\in\mathbb{N}}$, up to the (possibly empty) initial piece $\overline{y}^{(0)} := (y^{(1)}, \dots, y^{(t_1-1)})$, from $(\widehat{y}^{(i)})_{i\in\mathbb{N}}$ by simply writing out the constituent words of each $\widehat{y}^{(i)}$ in the correct order. Thus, $(y^{(n)})_{n\in\mathbb{N}} \leftrightarrow \big(\overline{y}^{(0)}, (\widehat{y}^{(i)})_{i\in\mathbb{N}}\big)$ is a 1-1-correspondence.

For $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ with $Q\big(T_1(Y) < \infty\big) = 1$ and $Y = (Y^{(n)})_{n\in\mathbb{N}} \sim Q$, we write

$$\widehat{Y}^{(i)} = \big(Y^{(T_i)}, Y^{(T_i+1)}, \dots, Y^{(T_{i+1}-1)}\big), \qquad i \in \mathbb{N}, \tag{2.15}$$

and call $\widehat{Y}^{(i)}$ the i-th excursion (note that $Y^{(T_{i+1})}$ is not included). We write $(\overline{Y}^{(0)}, (\widehat{Y}^{(i)})_{i\in\mathbb{N}})$ for what results after applying the above operation to Y. In ergodic-theoretic terms, $\widehat{Y} = (\widehat{Y}^{(i)})_{i\in\mathbb{N}}$ is the induced process corresponding to Y and the base set $E^{\geq \mathrm{tr}} \subseteq \widetilde{E}$ (see e.g. [S96, Section I.2.c] or [CFS82, Chapter 1, §5]). By construction, the sequence $\widehat{Y} = (\widehat{Y}^{(i)})_{i\in\mathbb{N}}$ is shift-invariant, i.e., $(\widehat{Y}^{(i)})_{i\in\mathbb{N}} =^d (\widehat{Y}^{(i+1)})_{i\in\mathbb{N}}$. We denote its law by $\widehat{Q}^{(\mathrm{tr})} \in \mathcal{P}^{\mathrm{inv}}\big((\widehat{E}^{(\mathrm{tr})})^{\mathbb{N}}\big)$. Note that the $\widetilde{E}^{\mathbb{N}}$-valued process obtained by writing out the constituent words of the excursions $\widehat{Y}^{(i)}$, $i \in \mathbb{N}$, in their correct order is in general not stationary. However, it is asymptotically mean stationary with stationary mean Q.

We also define $\widehat{\varrho}_{\geq \mathrm{tr}} \colon \cup_{N\in\mathbb{N}} \widetilde{E}^N \to \bigcup_{M\in\mathbb{N}_0} (\widehat{E}^{(\mathrm{tr})})^M$ as the operation on finite sequences of words as defined in (2.14), where we simply cut off a possible trailing end: For $y \in \widetilde{E}^N$, we let i in (2.14) only run up to $\#\{1 \leq m \leq N \colon |y^{(m)}| \geq \mathrm{tr}\} - 1$, which will be needed in Section 2.4 and in Appendix B.

Note that, in (2.15), $T_i = T_i(Y)$ is the index of the i-th occurrence of a word of length $\geq \mathrm{tr}$ in the sequence $Y = (Y^{(j)})_{j\in\mathbb{N}}$, which is drawn from Q. This is a small notational clash with (1.2), where the same letter refers to the renewal time points that $\mathbb{P}(\,\cdot \mid X)$ averages over. However, the meaning of the T_i's will always be clear from the context. In fact, in this section we will not use (1.2) explicitly and T_i will always refer to the i-th return time of Y to $E^{\geq \mathrm{tr}}$ as in (2.15).

Lemma 2.5. [Return time] *For $Q \in \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})$ and $\mathrm{tr} \in \mathbb{N}$, $\widehat{Q}^{(\mathrm{tr})}$ is ergodic w.r.t. the shift on $(\widehat{E}^{(\mathrm{tr})})^{\mathbb{N}}$, and*

$$\mathbb{E}_{\widehat{Q}^{(\mathrm{tr})}}\big[\mathsf{len}(\widehat{Y}^{(i)})\big] = \frac{1}{\varepsilon_{\mathrm{tr}}}, \tag{2.16}$$

where $\varepsilon_{\mathrm{tr}} = \varepsilon_{\mathrm{tr}}(Q) = Q(\tau_1 \geq \mathrm{tr})$.

Proof. As noted above, $\widehat{Y}$ in (2.15) is the induced process of Y observed when Y returns to $E^{\geq \mathrm{tr}}$. Note that $Q(Y^{(1)} \in E^{\geq \mathrm{tr}}) = \varepsilon_{\mathrm{tr}} > 0$, so $\widehat{Y}$ is well-defined and ergodic (see [S96, Theorem. I.2.19] or [CFS82, Chapter 1, §5, Theorem 1]). Equation (2.16) is the Kac formula for the expected return time to $E^{\geq \mathrm{tr}}$ under Q (see e.g. [S96, Eq. (14) in Section I.2.c] or [CFS82, Chapter 1, §5, Corollary to Lemma 1]). □

Lemma 2.6. [Variation on Abramov formula] *For every $\mathrm{tr} \in \mathbb{N}$,*

$$\begin{aligned} H(Q) &= \varepsilon_{\mathrm{tr}} H\big(\widehat{Q}^{(\mathrm{tr})}\big), \\ H(Q \mid Q_{\rho,\nu}) &= \varepsilon_{\mathrm{tr}} H\big(\widehat{Q}^{(\mathrm{tr})} \,\big|\, \widehat{Q}^{(\mathrm{tr})}_{\rho,\nu}\big). \end{aligned} \tag{2.17}$$

Moreover,

$$H([Q]_{\mathrm{tr}'}) = \varepsilon_{\mathrm{tr}} H\Big(\widehat{[Q]_{\mathrm{tr}'}}^{(\mathrm{tr})}\Big) \quad \forall\, \mathrm{tr}' \geq \mathrm{tr}. \tag{2.18}$$

Proof. See Appendix B. □

2.3.2 Glueing words inside an excursion together

With $\widehat{Y} = (\widehat{Y}^{(i)})_{i\in\mathbb{N}}$ from (2.15), write $\kappa(\widehat{Y}^{(i)}) = \kappa(Y^{(T_i)}, Y^{(T_i+1)}, \dots, Y^{(T_{i+1}-1)})$ for the concatenation of the words inside the i-th excursion, and denote the law of $(\kappa(\widehat{Y}^{(i)}))_{i\in\mathbb{N}}$ by $\widehat{\kappa}_{\mathrm{tr}}(Q)$. To keep track of the information required to invert this concatenation, put

$$\widehat{\tau}^{(i)} = \big(|Y^{(T_i+1)}|, |Y^{(T_i+2)}|, \dots, |Y^{(T_{i+1}-1)}|\big), \qquad i \in \mathbb{N}. \tag{2.19}$$

Each $\widehat{\tau}^{(i)}$ takes values in $\cup_{m\in\mathbb{N}_0}\mathbb{N}^m$ (we interpret $\mathbb{N}^0$ as consisting of only the empty sequence), and is by construction a composition of $|\kappa(\widehat{Y}^{(i)})| - |Y^{(T_i)}|$, i.e., its entries sum to $|\kappa(\widehat{Y}^{(i)})| - |Y^{(T_i)}|$. When $Y \sim [Q]_{\mathrm{tr}}$, the length of the first word $Y^{(T_i)}$ in $\widehat{Y}^{(i)}$ is exactly $= \mathrm{tr}$ by construction, which is why we do not keep track of its length in this notation. Therefore, indeed, $\widehat{Y}^{(i)} \leftrightarrow (\kappa(\widehat{Y}^{(i)}), \widehat{\tau}^{(i)})$ is a 1-1-correspondence.

Lemma 2.7. [Letters in the excursion decomposition]
For every $\mathrm{tr} \in \mathbb{N}$, $\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}}) \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$, $\kappa(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})) \in \mathcal{P}(E^{\mathbb{N}})$ *is asymptotically mean stationary with stationary mean* $\overline{\kappa\big(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big)} = \overline{\kappa([Q]_{\mathrm{tr}})} = \Psi_{[Q]_{\mathrm{tr}}}$, *and*

$$m_{[Q]_{\mathrm{tr}}} = \varepsilon_{\mathrm{tr}} m_{\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})}. \tag{2.20}$$

In particular,

$$\begin{aligned} m_{[Q]_{\mathrm{tr}}} H\big(\Psi_{[Q]_{\mathrm{tr}}}\big) &= \varepsilon_{\mathrm{tr}} m_{\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})} H\big(\kappa(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}}))\big), \\ m_{[Q]_{\mathrm{tr}}} H\big(\Psi_{[Q]_{\mathrm{tr}}} \,\big|\, \nu^{\otimes\mathbb{N}}\big) &= \varepsilon_{\mathrm{tr}} m_{\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})} H\big(\kappa(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})) \,\big|\, \nu^{\otimes\mathbb{N}}\big). \end{aligned} \tag{2.21}$$

Proof. For (2.20), note that $N \gg 1$ words from $[Q]_{\mathrm{tr}}$ correspond to $\varepsilon_{\mathrm{tr}} N + o(N)$ excursions from $E^{\geq \mathrm{tr}}$, and hence to $\varepsilon_{\mathrm{tr}} N + o(N)$ words from $\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})$. See Appendix B.3 for details. □

Lemma 2.8. [Concatenated excursions have small conditional entropy of the word lengths given their concatenation] *As* $\mathrm{tr} \to \infty$,

$$H\big(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) = m_{\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})} H\big(\kappa(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}}))\big) + O\big(\log(1/\varepsilon_{\mathrm{tr}}) + \log \mathrm{tr}\big), \tag{2.22}$$

where $\log(1/\varepsilon_{\mathrm{tr}}) + \log \mathrm{tr} = o(1/\varepsilon_{\mathrm{tr}})$.

Proof. Apply [B08, Lemma 3] to $\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})$, to obtain

$$m_{\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})} H\big(\kappa(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}}))\big) = H\big(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) - H_{\tau|K}\big(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big). \tag{2.23}$$

Further note that

$$0 \leq H_{\tau|K}\big(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) \leq H\Big(\mathscr{L}_{\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})}\big((|Y^{(i)}|)_{i\in\mathbb{N}}\big)\Big) \leq h\big(\mathscr{L}_{\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})}(|Y^{(1)}|)\big). \tag{2.24}$$

The first inequality follows from Lemma A.1, the second inequality is a general fact about specific entropy. Now estimate

$$\begin{aligned} h\big(\mathscr{L}_{\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})}(|Y^{(1)}|)\big) &\leq c_1 + c_2 \mathbb{E}_{\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})}\big[\log|Y^{(1)}|\big] \\ &\leq c_1 + c_2 \log \mathbb{E}_{\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})}\big[|Y^{(1)}|\big] = c_1 + c_2 \log\Big(\frac{1}{\varepsilon_{\mathrm{tr}}} m_{[Q]_{\mathrm{tr}}}\Big), \end{aligned} \tag{2.25}$$

where we use Lemma A.3 in the first inequality, Jensen's inequality in the second inequality, and (2.20) from Lemma 2.7 for the last equality. Note that

$$\varepsilon_{\mathrm{tr}} \log(m_{[Q]_{\mathrm{tr}}}) \leq \varepsilon_{\mathrm{tr}} \log(\mathrm{tr}) \leq \mathbb{E}_Q\big[\mathbb{1}_{\{|Y^{(1)}|\geq \mathrm{tr}\}} \log(|Y^{(1)}|)\big] \underset{\mathrm{tr}\to\infty}{\downarrow} 0 \tag{2.26}$$

because $\mathbb{E}_Q\big[\log(|Y^{(1)}|)\big] < \infty$ by Lemma A.2. □

Below, for a probability measure P and a discrete random variable U, we will write $P(U)$ for the random variable $f(U)$, where $f(u) = P(U = u)$. Analogously, $P(U \,|\, V)$ means $g(U, V)$, where $g(u, v) = P(U = u \,|\, V = v)$.

Lemma 2.9. [Entropy decomposition for the induced process] *For every* $\mathrm{tr} \in \mathbb{N}$,

$$H\big(\widehat{[Q]_{\mathrm{tr}}}^{(\mathrm{tr})}\big) = H\big(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) + H_{\widehat{\tau}|\widehat{\kappa}_{\mathrm{tr}}}\big([Q]_{\mathrm{tr}}\big), \tag{2.27}$$

where

$$H_{\widehat{\tau}|\widehat{\kappa}_{\mathrm{tr}}}\big([Q]_{\mathrm{tr}}\big) = \lim_{N\to\infty} \mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[-\frac{1}{N} \log [Q]_{\mathrm{tr}}\big(\widehat{\tau}^{(1)}, \dots, \widehat{\tau}^{(N)} \,|\, \kappa(\widehat{Y}^{(1)}), \dots, \kappa(\widehat{Y}^{(N)})\big)\Big] \tag{2.28}$$

is the specific conditional entropy of the word lengths inside the excursions given their concatenations under the law $[Q]_{\mathrm{tr}}$.

Proof. Write

$$\begin{aligned}
[Q]_{\mathrm{tr}}\big(\widehat{Y}^{(1)}, \dots \widehat{Y}^{(N)}\big) &= [Q]_{\mathrm{tr}}\big(\kappa(\widehat{Y}^{(1)}), \dots, \kappa(\widehat{Y}^{(N)})\big)\, [Q]_{\mathrm{tr}}\big(\widehat{Y}^{(1)}, \dots \widehat{Y}^{(N)} \,|\, \kappa(\widehat{Y}^{(1)}), \dots, \kappa(\widehat{Y}^{(N)})\big) \\
&= [Q]_{\mathrm{tr}}\big(\kappa(\widehat{Y}^{(1)}), \dots, \kappa(\widehat{Y}^{(N)})\big)\, [Q]_{\mathrm{tr}}\big(\widehat{\tau}^{(1)}, \dots \widehat{\tau}^{(N)} \,|\, \kappa(\widehat{Y}^{(1)}), \dots, \kappa(\widehat{Y}^{(N)})\big),
\end{aligned} \tag{2.29}$$

take the logarithm, divide by $-N$, and use ergodicity and the Shannon-McMillan-Breiman theorem. □

Combining (2.27) with (2.18) from Lemma 2.6 for $\mathrm{tr}' = \mathrm{tr}$, we get

$$H\big([Q]_{\mathrm{tr}}\big) = \varepsilon_{\mathrm{tr}} H\big(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) + \varepsilon_{\mathrm{tr}} H_{\widehat{\tau}|\widehat{\kappa}_{\mathrm{tr}}}\big([Q]_{\mathrm{tr}}\big). \tag{2.30}$$

2.3.3 Removing the content of long words

We next define $\varphi_{\mathrm{tr}} \colon \widetilde{E} \to \widetilde{E}_{\mathrm{tr}} := \big(\cup_{n=1}^{\mathrm{tr}-1} E^n\big) \cup \{*_{\geq \mathrm{tr}}\}$ by leaving short words unchanged and replacing long words by a special symbol $*_{\geq \mathrm{tr}} \notin E \cup \widetilde{E}$, i.e.,

$$\varphi_{\mathrm{tr}}(w) = \begin{cases} w, & |w| < \mathrm{tr}, \\ *_{\geq \mathrm{tr}}, & |w| \geq \mathrm{tr}. \end{cases} \tag{2.31}$$

Extend φ_{tr} in the obvious way to mappings $\varphi_{\mathrm{tr}} \colon \widetilde{E}^N \to \widetilde{E}_{\mathrm{tr}}^N$, $N \in \mathbb{N}$, $\varphi_{\mathrm{tr}} \colon \widetilde{E}^{\mathbb{N}} \to \widetilde{E}_{\mathrm{tr}}^{\mathbb{N}}$ and $\varphi_{\mathrm{tr}} \colon \mathcal{P}(\widetilde{E}^{\mathbb{N}}) \to \mathcal{P}\big(\widetilde{E}_{\mathrm{tr}}^{\mathbb{N}}\big)$ by applying φ_{tr} coordinate-wise and taking the image measure, respectively. By construction, $\varphi_{\mathrm{tr}} \colon \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) \to \mathcal{P}^{\mathrm{inv}}\big(\widetilde{E}_{\mathrm{tr}}^{\mathbb{N}}\big)$ and $\varphi_{\mathrm{tr}}([Q]_{\mathrm{tr}}) = \varphi_{\mathrm{tr}}(Q)$.

Lemma 2.10. [Approximating specific relative entropy by removing the content of long words completely] *For every* $\mathrm{tr} \in \mathbb{N}$, $\varphi_{\mathrm{tr}}(Q) \in \mathcal{P}^{\mathrm{erg}}\big(\widetilde{E}_{\mathrm{tr}}^{\mathbb{N}}\big)$,

$$H\big(\varphi_{\mathrm{tr}}(Q) \,\big|\, \varphi_{\mathrm{tr}}(Q_{\rho,\nu})\big) \leq H\big([Q]_{\mathrm{tr}} \,\big|\, [Q_{\rho,\nu}]_{\mathrm{tr}}\big) \leq H(Q \mid Q_{\rho,\nu}), \tag{2.32}$$

while

$$\lim_{\mathrm{tr}\to\infty} H\big(\varphi_{\mathrm{tr}}(Q) \,\big|\, \varphi_{\mathrm{tr}}(Q_{\rho,\nu})\big) = H(Q \mid Q_{\rho,\nu}). \tag{2.33}$$

Proof. Note that $\varphi_{\mathrm{tr}}(Q)$ is a stationary coding of Q and hence inherits ergodicity from Q. The inequalities in (2.32) follow from the fact that projections can only decrease relative entropy. To check (2.33), observe that, for any $N \in \mathbb{N}$ and any $Q \in \mathcal{P}(\widetilde{E}^{\mathbb{N}})$, $\varphi_{\mathrm{tr}}(\pi_N Q) = \pi_N \varphi_{\mathrm{tr}}(Q)$, and

$$h\big(\pi_N Q \,\big|\, \pi_N Q_{\rho,\nu}\big) = \lim_{\mathrm{tr}\to\infty} h\big(\varphi_{\mathrm{tr}}(\pi_N Q) \,\big|\, \varphi_{\mathrm{tr}}(\pi_N Q_{\rho,\nu})\big) = \sup_{\mathrm{tr}\in\mathbb{N}} h\big(\varphi_{\mathrm{tr}}(\pi_N Q) \,\big|\, \varphi_{\mathrm{tr}}(\pi_N Q_{\rho,\nu})\big). \tag{2.34}$$

Consequently,

$$\begin{aligned} H(Q \mid Q_{\rho,\nu}) &= \sup_{N\in\mathbb{N}} \frac{1}{N} h\big(\pi_N Q \,\big|\, \pi_N Q_{\rho,\nu}\big) = \sup_{N\in\mathbb{N}} \sup_{\mathrm{tr}\in\mathbb{N}} \frac{1}{N} h\big(\varphi_{\mathrm{tr}}(\pi_N Q) \,\big|\, \varphi_{\mathrm{tr}}(\pi_N Q_{\rho,\nu})\big) \\ &= \sup_{\mathrm{tr}\in\mathbb{N}} \sup_{N\in\mathbb{N}} \frac{1}{N} h\big(\varphi_{\mathrm{tr}}(\pi_N Q) \,\big|\, \varphi_{\mathrm{tr}}(\pi_N Q_{\rho,\nu})\big) = \sup_{\mathrm{tr}\in\mathbb{N}} \sup_{N\in\mathbb{N}} \frac{1}{N} h\big(\pi_N \varphi_{\mathrm{tr}}(Q) \,\big|\, \pi_N \varphi_{\mathrm{tr}}(Q_{\rho,\nu})\big) \\ &= \sup_{\mathrm{tr}\in\mathbb{N}} H\big(\varphi_{\mathrm{tr}}(Q) \,\big|\, \varphi_{\mathrm{tr}}(Q_{\rho,\nu})\big) = \lim_{\mathrm{tr}\to\infty} H\big(\varphi_{\mathrm{tr}}(Q) \,\big|\, \varphi_{\mathrm{tr}}(Q_{\rho,\nu})\big). \end{aligned} \tag{2.35}$$

□

Corollary 2.11. [Control of the specific entropy drop upon removal of long words] *As* $\mathrm{tr} \to \infty$,

$$H\big([Q]_{\mathrm{tr}} \,\big|\, [Q_{\rho,\nu}]_{\mathrm{tr}}\big) - H\big(\varphi_{\mathrm{tr}}(Q) \,\big|\, \varphi_{\mathrm{tr}}(Q_{\rho,\nu})\big) = o(1) \tag{2.36}$$

and

$$H\big([Q]_{\mathrm{tr}}\big) - H\big(\varphi_{\mathrm{tr}}(Q)\big) = -\mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[\mathbb{1}_{\{|Y^{(1)}|=\mathrm{tr}\}} \sum_{i=1}^{|Y^{(1)}|} \log \nu(Y_i^{(1)})\Big] + o(1). \tag{2.37}$$

Proof. The claim in (2.36) is an immediate consequence of (2.32)–(2.33) in Lemma 2.10. For the proof of (2.37), note that

$$\begin{aligned} &H\big([Q]_{\mathrm{tr}} \,\big|\, [Q_{\rho,\nu}]_{\mathrm{tr}}\big) - H\big(\varphi_{\mathrm{tr}}(Q) \,\big|\, \varphi_{\mathrm{tr}}(Q_{\rho,\nu})\big) \\ &= -H\big([Q]_{\mathrm{tr}}\big) - \mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[\mathbb{1}_{\{|Y^{(1)}|<\mathrm{tr}\}} \log \rho(|Y^{(1)}|) + \mathbb{1}_{\{|Y^{(1)}|=\mathrm{tr}\}} \log \rho([\mathrm{tr},\infty)) + \textstyle\sum_{i=1}^{|Y^{(1)}|} \log \nu(Y_i^{(1)})\Big] \\ &\quad + H\big(\varphi_{\mathrm{tr}}(Q)\big) \\ &\quad + \mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[\mathbb{1}_{\{|Y^{(1)}|<\mathrm{tr}\}}\Big(\log \rho(|Y^{(1)}|) + \textstyle\sum_{i=1}^{|Y^{(1)}|} \log \nu(Y_i^{(1)})\Big) + \mathbb{1}_{\{|Y^{(1)}|=\mathrm{tr}\}} \log \rho([\mathrm{tr},\infty))\Big] \\ &= -H\big([Q]_{\mathrm{tr}}\big) + H\big(\varphi_{\mathrm{tr}}(Q)\big) - \mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[\mathbb{1}_{\{|Y^{(1)}|=\mathrm{tr}\}} \sum_{i=1}^{|Y^{(1)}|} \log \nu(Y_i^{(1)})\Big]. \end{aligned} \tag{2.38}$$

□

2.3.4 Excursion picture with long words removed

Define $\widehat{\kappa}^*_{\mathrm{tr}}\colon \widetilde{E}^{\mathbb{N}} \to (E^*)^{\mathbb{N}}$, where $E^* = \cup_{m\in\mathbb{N}_0} E^m$ and $E^0 = \{\epsilon\}$ is the set consisting of exactly the empty word ϵ, as follows. Write

$$\pi_{-1}\widehat{Y}^{(i)} = \big(Y^{(T_i+1)}, Y^{(T_i+2)}, \ldots, Y^{(T_{i+1}-1)}\big) \tag{2.39}$$

for the i-th excursion without its first word, and

$$\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(i)}) = \kappa\big(\pi_{-1}\widehat{Y}^{(i)}\big) \in E^* \tag{2.40}$$

for the (possibly empty) concatenation of the short words in the i-th excursion. Note that $\pi_{-1}Y^{(i)} \leftrightarrow (\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(i)}), \widehat{\tau}^{(i)})$ is a 1-1-correspondence. When $Y \sim Q$, we denote the law of $\big(\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(i)})\big)_{i\in\mathbb{N}}$ by $\widehat{\kappa}^*_{\mathrm{tr}}(Q)$. Note that $\widehat{\kappa}^*_{\mathrm{tr}}(Q) = \widehat{\kappa}^*_{\mathrm{tr}}([Q]_{\mathrm{tr}})$ by construction.

Lemma 2.12. [Entropy decomposition according to excursions with long words removed] *For every* $\mathrm{tr} \in \mathbb{N}$, $\widehat{\kappa}^*_{\mathrm{tr}}([Q]_{\mathrm{tr}}) \in \mathcal{P}^{\mathrm{erg,fin}}\big((E^*)^{\mathbb{N}}\big)$,

$$H\big(\varphi_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) = \varepsilon_{\mathrm{tr}} H\big(\widehat{\kappa}^*_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) + \varepsilon_{\mathrm{tr}} H_{\widehat{\tau}|\widehat{\kappa}^*_{\mathrm{tr}}}\big([Q]_{\mathrm{tr}}\big), \tag{2.41}$$

where

$$H_{\widehat{\tau}|\widehat{\kappa}^*_{\mathrm{tr}}}\big([Q]_{\mathrm{tr}}\big) = \lim_{N\to\infty} \mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[-\frac{1}{N}\log [Q]_{\mathrm{tr}}\big(\widehat{\tau}^{(1)},\dots \widehat{\tau}^{(N)} \,|\, \widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(1)}),\dots,\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(N)})\big)\Big] \tag{2.42}$$

is the specific conditional entropy of the word lengths inside the excursions given the concatenations of their short words under the law $[Q]_{\mathrm{tr}}$.

Proof. Apply the arguments from Lemma 2.9 and Lemma 2.6 to $\varphi_{\mathrm{tr}}([Q]_{\mathrm{tr}})$ instead of $[Q]_{\mathrm{tr}}$. The fact that in $\varphi_{\mathrm{tr}}([Q]_{\mathrm{tr}})$ the content of the long words is collapsed into the special symbol $*_{\geq \mathrm{tr}}$ requires only notational adaptations. □

2.3.5 Quantifications

Lemma 2.13. [Conditional on the short words the long words are close to i.i.d. draws from ν] *As* $\mathrm{tr} \to \infty$,

$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[h\Big(\mathscr{L}\Big(\big(\widehat{Y}^{(1)}_1, \widehat{Y}^{(2)}_1,\dots,\widehat{Y}^{(N)}_1\big) \,\Big|\, \big(\pi_{-1}\widehat{Y}^{(1)}, \pi_{-1}\widehat{Y}^{(2)},\dots,\pi_{-1}\widehat{Y}^{(N)}\big)\Big) \,\Big|\, \big(\nu^{\otimes \mathrm{tr}}\big)^{\otimes N}\Big)\Big] \\
&\quad = -\frac{1}{\varepsilon_{\mathrm{tr}}}\left(H\big([Q]_{\mathrm{tr}}\big) - H\big(\varphi_{\mathrm{tr}}(Q)\big) + \mathbb{E}_{[Q]_{\mathrm{tr}}}\Bigg[\mathbb{1}_{\{|Y^{(1)}|=\mathrm{tr}\}} \sum_{i=1}^{|Y^{(1)}|} \log \nu(Y^{(1)}_i)\Bigg]\right) = o\Big(\frac{1}{\varepsilon_{\mathrm{tr}}}\Big).
\end{aligned} \tag{2.43}$$

Proof. (In Remark A.8 we provide a quick intuition of this estimate by performing a one-marginal computation.) We can interpret (2.37) from Corollary 2.11 as follows. By ergodicity of $[Q]_{\mathrm{tr}}$ and $\varphi_{\mathrm{tr}}(Q)$ (inherited from ergodicity of Q) and the Shannon-McMillan-Breiman theorem (see [S96, Theorem I.5.1] or [B65]), we have

$$\lim_{N\to\infty} \frac{1}{N}\log[Q]_{\mathrm{tr}}\big(Y^{(1,\dots,N)} \,\big|\, \varphi_{\mathrm{tr}}(Y^{(1,\dots,N)})\big) = H\big([Q]_{\mathrm{tr}}\big) - H\big(\varphi_{\mathrm{tr}}(Q)\big) \quad [Q]_{\mathrm{tr}}\text{-a.s. and in } \mathcal{L}^1\big([Q]_{\mathrm{tr}}\big). \tag{2.44}$$

Think of this limit as the specific entropy of the long words conditional on the short words (write $[Q]_{\mathrm{tr}}(Y^{(1,\dots,N)}) = [Q]_{\mathrm{tr}}(\varphi_{\mathrm{tr}}(Y^{(1,\dots,N)}))\,[Q]_{\mathrm{tr}}(Y^{(1,\dots,N)} \,|\, \varphi_{\mathrm{tr}}(Y^{(1,\dots,N)}))$, etc.). Further note that $(\widehat{Y}^{(1)}_1, \widehat{Y}^{(2)}_1,\dots,\widehat{Y}^{(N)}_1)$ is a random variable with values in $(E^{\mathrm{tr}})^N$, and

$$\mathscr{L}\Big(\big(\widehat{Y}^{(1)}_1, \widehat{Y}^{(2)}_1,\dots,\widehat{Y}^{(N)}_1\big) \,\Big|\, \big(\pi_{-1}\widehat{Y}^{(1)}, \pi_{-1}\widehat{Y}^{(2)},\dots,\pi_{-1}\widehat{Y}^{(N)}\big)\Big) \tag{2.45}$$

is a (random) probability distribution on $(E^{\mathrm{tr}})^N$. We have

$$\mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[h\Big(\mathscr{L}\Big(\big(\widehat{Y}^{(1)}_1, \widehat{Y}^{(2)}_1,\dots,\widehat{Y}^{(N)}_1\big) \,\Big|\, \big(\pi_{-1}\widehat{Y}^{(1)}, \pi_{-1}\widehat{Y}^{(2)},\dots,\pi_{-1}\widehat{Y}^{(N)}\big)\Big) \,\Big|\, \big(\nu^{\otimes \mathrm{tr}}\big)^{\otimes N}\Big)\Big]$$

$$
\begin{aligned}
&= -\mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[h\Big(\mathscr{L}\Big(\big(\widehat{Y}_1^{(1)},\widehat{Y}_1^{(2)},\dots,\widehat{Y}_1^{(N)}\big)\,\Big|\,\big(\pi_{-1}\widehat{Y}^{(1)},\pi_{-1}\widehat{Y}^{(2)},\dots,\pi_{-1}\widehat{Y}^{(N)}\big)\Big)\Big)\Big] \\
&\qquad - N\mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[\sum_{j=1}^{\mathrm{tr}}\log\nu\big(\widehat{Y}_{1,j}^{(1)}\big)\Big],
\end{aligned} \tag{2.46}
$$

where $\widehat{Y}_{1,j}^{(1)}$ denotes the j-th letter in the word $\widehat{Y}_1^{(1)}$. Furthermore,

$$
\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[h\Big(\mathscr{L}\Big(\big(\widehat{Y}_1^{(1)},\widehat{Y}_1^{(2)},\dots,\widehat{Y}_1^{(N)}\big)\,\Big|\,\big(\pi_{-1}\widehat{Y}^{(1)},\pi_{-1}\widehat{Y}^{(2)},\dots,\pi_{-1}\widehat{Y}^{(N)}\big)\Big)\Big)\Big] \\
&\quad = \lim_{N\to\infty}\frac{1}{\varepsilon_{\mathrm{tr}}N}\mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[h\Big(\mathscr{L}\big(Y^{(1,\dots,N)}\,|\,\varphi_{\mathrm{tr}}(Y^{(1,\dots,N)})\big)\Big)\Big] = \frac{1}{\varepsilon_{\mathrm{tr}}}\Big(H\big([Q]_{\mathrm{tr}}\big) - H\big(\varphi_{\mathrm{tr}}(Q)\big)\Big),
\end{aligned} \tag{2.47}
$$

which is a "conditional Abramov formula" analogous to Lemma 2.6. Note that N consecutive words drawn from $[Q]_{\mathrm{tr}}$ contain approximately $\varepsilon_{\mathrm{tr}}N$ many excursions (see Appendix B for more details). Finally, we have

$$
\mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[\sum_{j=1}^{\mathrm{tr}}\log\nu\big(\widehat{Y}_{1,j}^{(1)}\big)\Big] = \frac{1}{\varepsilon_{\mathrm{tr}}}\mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[\mathbb{1}_{\{|Y^{(1)}|=\mathrm{tr}\}}\sum_{i=1}^{|Y^{(1)}|}\log\nu\big(Y_i^{(1)}\big)\Big]. \tag{2.48}
$$

Thus, (2.43) follows from (2.37). □

Lemma 2.14. [Negligible contribution of the long words to the conditional entropy of word lengths] *As* $\mathrm{tr}\to\infty$,

$$
0 \le H_{\widehat{\tau}|\widehat{\kappa}^*_{\mathrm{tr}}}\big([Q]_{\mathrm{tr}}\big) - H_{\widehat{\tau}|\widehat{\kappa}_{\mathrm{tr}}}\big([Q]_{\mathrm{tr}}\big) = o\Big(\frac{1}{\varepsilon_{\mathrm{tr}}}\Big). \tag{2.49}
$$

Proof. We have

$$
\begin{aligned}
&-\log\frac{[Q]_{\mathrm{tr}}\big(\pi_{-1}\widehat{Y}^{(1)},\dots,\pi_{-1}\widehat{Y}^{(N)}\big)}{[Q]_{\mathrm{tr}}\big(\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(1)}),\dots,\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(N)})\big)} + \log\frac{[Q]_{\mathrm{tr}}\big(\widehat{Y}^{(1)},\dots,\widehat{Y}^{(N)}\big)}{[Q]_{\mathrm{tr}}\big(\kappa(\widehat{Y}^{(1)}),\dots,\kappa(\widehat{Y}^{(N)})\big)} \\
&\quad = \log\frac{[Q]_{\mathrm{tr}}\big(\widehat{Y}^{(1)},\dots,\widehat{Y}^{(N)}\big)}{[Q]_{\mathrm{tr}}\big(\pi_{-1}\widehat{Y}^{(1)},\dots,\pi_{-1}\widehat{Y}^{(N)}\big)} - \log\frac{[Q]_{\mathrm{tr}}\big(\kappa(\widehat{Y}^{(1)}),\dots,\kappa(\widehat{Y}^{(N)})\big)}{[Q]_{\mathrm{tr}}\big(\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(1)}),\dots,\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(N)})\big)} \\
&\quad = \log[Q]_{\mathrm{tr}}\big(Y^{(T_1)},\dots,Y^{(T_N)}\,|\,\pi_{-1}\widehat{Y}^{(1)},\dots,\pi_{-1}\widehat{Y}^{(N)}\big) \\
&\qquad - \log[Q]_{\mathrm{tr}}\big(Y^{(T_1)},\dots,Y^{(T_N)}\,|\,\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(1)}),\dots,\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(N)})\big)
\end{aligned} \tag{2.50}
$$

(note that $\pi_{-1}\widehat{Y}^{(i)} \leftrightarrow \big(\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(i)}),\widehat{\tau}^{(i)}\big)$, $\widehat{Y}^{(i)} \leftrightarrow \big(\kappa(\widehat{Y}^{(i)}),\widehat{\tau}^{(i)}\big)$ and $\kappa(\widehat{Y}^{(i)}) \leftrightarrow \big(Y^{(T_i)},\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(i)})\big)$ are all 1-1-encodings when Y is drawn from $[Q]_{\mathrm{tr}}$). Thus, using the Shannon-McMillan-Breiman theorem (see [S96, Thoerem I.5.1] or [B65]), we can write

$$
\begin{aligned}
&H_{\widehat{\tau}|\widehat{\kappa}^*_{\mathrm{tr}}}\big([Q]_{\mathrm{tr}}\big) - H_{\widehat{\tau}|\widehat{\kappa}_{\mathrm{tr}}}\big([Q]_{\mathrm{tr}}\big) \\
&\quad = \lim_{N\to\infty}\mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[-\frac{1}{N}\log[Q]_{\mathrm{tr}}\big(Y^{(T_1)},\dots,Y^{(T_N)}\,|\,\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(1)}),\dots,\widehat{\kappa}^*_{\mathrm{tr}}(\widehat{Y}^{(N)})\big)\Big] \\
&\qquad - \lim_{N\to\infty}\mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[-\frac{1}{N}\log[Q]_{\mathrm{tr}}\big(Y^{(T_1)},\dots,Y^{(T_N)}\,|\,\pi_{-1}\widehat{Y}^{(1)},\dots,\pi_{-1}\widehat{Y}^{(N)}\big)\Big]
\end{aligned}
$$

$$= \lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{[Q]_{\rm tr}}\Big[h\Big(\mathscr{L}_{[Q]_{\rm tr}}\big(Y^{(T_1)},\dots,Y^{(T_N)}\,|\,\widehat{\kappa}^*_{\rm tr}(\widehat{Y}^{(1)}),\dots,\widehat{\kappa}^*_{\rm tr}(\widehat{Y}^{(N)})\big)\Big)\Big]$$
$$- \lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{[Q]_{\rm tr}}\Big[h\Big(\mathscr{L}_{[Q]_{\rm tr}}\big(Y^{(T_1)},\dots,Y^{(T_N)}\,|\,\pi_{-1}\widehat{Y}^{(1)},\dots,\pi_{-1}\widehat{Y}^{(N)}\big)\Big)\Big] \ge 0. \tag{2.51}$$

The last inequality follows from the general fact that conditioning on a smaller σ-algebra can only increase the expected entropy of a conditional distribution (see Observation A.5 below). For quantitative control on the difference in (2.49), note that (2.43) from Lemma 2.13 shows that the expected specific conditional relative entropy of the content of the long words w.r.t. $\nu^{\otimes \rm tr}$ is suitably controlled, i.e., is $o(1/\varepsilon_{\rm tr})$), and use Observation A.7 below. □

Proposition 2.15. [Approximate representations of entropy] *As* tr $\to \infty$,

$$m_{[Q]_{\rm tr}} H\big(\Psi_{[Q]_{\rm tr}}\,\big|\,\nu^{\otimes\mathbb{N}}\big) = -\varepsilon_{\rm tr} H\big(\widehat{\kappa}^*_{\rm tr}([Q]_{\rm tr})\big) - \mathbb{E}_{[Q]_{\rm tr}}\Big[\mathbb{1}_{\{|Y^{(1)}|<{\rm tr}\}}\sum_{j=1}^{|Y^{(1)}|}\log\nu(Y^{(1)}_j)\Big] + o(1), \tag{2.52}$$
$$H_{\tau|K}([Q]_{\rm tr}) = \varepsilon_{\rm tr} H_{\widehat{\tau}|\widehat{\kappa}^*_{\rm tr}}\big([Q]_{\rm tr}\big) + o(1). \tag{2.53}$$

Proof. For the proof of (2.52), note that

$$H\big(\kappa(\widehat{\kappa}_{\rm tr}([Q]_{\rm tr}))\,\big|\,\nu^{\otimes\mathbb{N}}\big) = -H\big(\kappa(\widehat{\kappa}_{\rm tr}([Q]_{\rm tr}))\big) - \lim_{L\to\infty}\frac{1}{L}\mathbb{E}_{\kappa(\widehat{\kappa}_{\rm tr}([Q]_{\rm tr}))}\Big[\sum_{j=1}^{L}\log\nu(X_j)\Big] \tag{2.54}$$

and

$$\begin{aligned}&\lim_{L\to\infty}\frac{1}{L}\mathbb{E}_{\kappa(\widehat{\kappa}_{\rm tr}([Q]_{\rm tr}))}\Big[\sum_{j=1}^{L}\log\nu(X_j)\Big]\\ &\quad= \lim_{L\to\infty}\frac{1}{L}\mathbb{E}_{\kappa([Q]_{\rm tr})}\Big[\sum_{j=1}^{L}\log\nu(X_j)\Big] = \frac{1}{m_{[Q]_{\rm tr}}}\mathbb{E}_{[Q]_{\rm tr}}\Big[\sum_{j=1}^{|Y^{(1)}|}\log\nu(Y^{(1)}_j)\Big],\end{aligned} \tag{2.55}$$

where we use $\overline{\kappa(\widehat{\kappa}_{\rm tr}([Q]_{\rm tr}))} = \overline{\kappa([Q]_{\rm tr})} = \Psi_{[Q]_{\rm tr}}$ from Lemma 2.7 for the first equality, and argue as in [B08, Section 3] for the second equality. Thus,

$$\begin{aligned} m_{[Q]_{\rm tr}} H\big(\Psi_{[Q]_{\rm tr}}\,\big|\,\nu^{\otimes\mathbb{N}}\big) &= \varepsilon_{\rm tr} m_{\widehat{\kappa}_{\rm tr}([Q]_{\rm tr})} H\big(\kappa(\widehat{\kappa}_{\rm tr}([Q]_{\rm tr}))\,\big|\,\nu^{\otimes\mathbb{N}}\big)\\ &= \varepsilon_{\rm tr} m_{\widehat{\kappa}_{\rm tr}([Q]_{\rm tr})}\Big(-H\big(\kappa(\widehat{\kappa}_{\rm tr}([Q]_{\rm tr}))\big) - \frac{1}{m_{[Q]_{\rm tr}}}\mathbb{E}_{[Q]_{\rm tr}}\Big[\sum_{j=1}^{|Y^{(1)}|}\log\nu(Y^{(1)}_j)\Big]\Big)\\ &= -\varepsilon_{\rm tr} H\big(\widehat{\kappa}_{\rm tr}([Q]_{\rm tr})\big) - \mathbb{E}_{[Q]_{\rm tr}}\Big[\sum_{j=1}^{|Y^{(1)}|}\log\nu(Y^{(1)}_j)\Big] + O\big(|\varepsilon_{\rm tr}\log\varepsilon_{\rm tr}| + \varepsilon_{\rm tr}\log{\rm tr}\big),\end{aligned} \tag{2.56}$$

where we use (2.22) in the last equation. Inserting (2.37), we get

$$\begin{aligned}& m_{[Q]_{\rm tr}} H\big(\kappa([Q]_{\rm tr})\,\big|\,\nu^{\otimes\mathbb{N}}\big)\\ &= -\varepsilon_{\rm tr} H\big(\widehat{\kappa}_{\rm tr}([Q]_{\rm tr})\big) - \mathbb{E}_{[Q]_{\rm tr}}\Big[\mathbb{1}_{\{|Y^{(1)}|<{\rm tr}\}}\sum_{j=1}^{|Y^{(1)}|}\log\nu(Y^{(i)}_j)\Big] + H\big([Q]_{\rm tr}\big) - H\big(\varphi_{\rm tr}(Q)\big) + o(1).\end{aligned} \tag{2.57}$$

Subtracting (2.30) and (2.41), and inserting (2.49) from Lemma 2.14, we find

$$
\begin{aligned}
&H\big([Q]_{\mathrm{tr}}\big) - H\big(\varphi_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) \\
&= \varepsilon_{\mathrm{tr}} H\big(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) - \varepsilon_{\mathrm{tr}} H\big(\widehat{\kappa}^*_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) + \varepsilon_{\mathrm{tr}}\Big(H_{\widehat{\tau}|\widehat{\kappa}_{\mathrm{tr}}}\big([Q]_{\mathrm{tr}}\big) - H_{\widehat{\tau}|\widehat{\kappa}^*_{\mathrm{tr}}}\big([Q]_{\mathrm{tr}}\big)\Big) \\
&= \varepsilon_{\mathrm{tr}} H\big(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) - \varepsilon_{\mathrm{tr}} H\big(\widehat{\kappa}^*_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) + o(1).
\end{aligned}
\tag{2.58}
$$

Combined with (2.57), this implies (2.52). To verify (2.53), write

$$
\begin{aligned}
\varepsilon_{\mathrm{tr}} H_{\widehat{\tau}|\widehat{\kappa}^*_{\mathrm{tr}}}\big([Q]_{\mathrm{tr}}\big) &= \varepsilon_{\mathrm{tr}} H_{\widehat{\tau}|\widehat{\kappa}_{\mathrm{tr}}}\big([Q]_{\mathrm{tr}}\big) + o(1) \\
&= H\big([Q]_{\mathrm{tr}}\big) - \varepsilon_{\mathrm{tr}} H\big(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) + o(1) \\
&= H\big([Q]_{\mathrm{tr}}\big) - \varepsilon_{\mathrm{tr}} m_{\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})} H\big(\kappa(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}}))\big) + o(1) \\
&= H\big([Q]_{\mathrm{tr}}\big) - m_{[Q]_{\mathrm{tr}}} H\big(\kappa([Q]_{\mathrm{tr}})\big) + o(1) \\
&= H\big([Q]_{\mathrm{tr}}\big) - m_{[Q]_{\mathrm{tr}}} H\big(\Psi_{[Q]_{\mathrm{tr}}}\big) + o(1) = H_{\tau|K}\big([Q]_{\mathrm{tr}}\big) + o(1).
\end{aligned}
\tag{2.59}
$$

Here, we use Lemma 2.14 for the first equality, (2.30) for the second equality, (2.22) for the third equality, (2.21) from Lemma 2.7 for the fourth equality, and [B08, Lemma 3] for the fifth equality (recall that $\Psi_{[Q]_{\mathrm{tr}}}$ is the asymptotically mean stationary measure of $\kappa([Q]_{\mathrm{tr}})$). □

2.4 Ergodicity and neighbourhoods

When we draw $Y \sim [Q]_{\mathrm{tr}}$, $\big(Y, \widehat{\kappa}_{\mathrm{tr}}(Y), \varphi_{\mathrm{tr}}(Y), \widehat{\kappa}^*_{\mathrm{tr}}(Y)\big)$ provides a natural coupling of $[Q]_{\mathrm{tr}}$, $\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})$, $\varphi_{\mathrm{tr}}([Q]_{\mathrm{tr}})$ and $\widehat{\kappa}^*_{\mathrm{tr}}([Q]_{\mathrm{tr}})$. Thus, we can augment the construction from [BGdH10, Section 3.1] and apply it to $[Q]_{\mathrm{tr}}$ as follows. Pick $\varepsilon_1 > 0$ and $\delta_1 > 0$ (we will let $\varepsilon_1 \downarrow 0$ and $\delta_1 \downarrow 0$ later). By ergodicity of $\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})$ and of $\varphi_{\mathrm{tr}}(Q)$, we can choose $M \in \mathbb{N}$ sufficiently large ($M \gg 1/\varepsilon_{\mathrm{tr}}$), and choose a finite set $\mathscr{A} \subset \widetilde{E}^M$ of $[Q]_{\mathrm{tr}}$-typical M-sentences with

$$
\pi_M [Q]_{\mathrm{tr}}(\mathscr{A}) > 1 - \delta_1 \tag{2.60}
$$

and the following properties: For every $y = (y^{(1)}, \dots, y^{(M)}) \in \mathscr{A}$ (cf. [BGdH10, Section 3.1, Eq. (3.6)])

$$
|\widehat{\kappa}^*_{\mathrm{tr}}(y)| \in \big[\varepsilon_{\mathrm{tr}} M \,\pm\, \varepsilon_1 \varepsilon_{\mathrm{tr}} M\big], \tag{2.61}
$$

$$
\sum_{\substack{i=1 \\ |y^{(i)}| < \mathrm{tr}}}^{M} \log \rho\big(|y^{(i)}|\big) \in \Big[M(1-\varepsilon_{\mathrm{tr}})\mathbb{E}_{[Q]_{\mathrm{tr}}}\big[\log \rho(\tau_1) \,\big|\, \tau_1 < \mathrm{tr}\big] \,\pm\, \varepsilon_1 M\Big]. \tag{2.62}
$$

(note that (2.60) ensures that (2.76) below indeed defines a neighbourhood of Q.) Here and below, we use the short-hand notations $[a \pm b] := [a-b, a+b]$ and similarly $[\exp(a \pm b)] := [\exp(a - b), \exp(a+b)]$. Note that, by Lemma 2.5, an excursion from $E^{\geq \mathrm{tr}}$ under Q (or, equivalently, under $[Q]_{\mathrm{tr}}$) contains on average $1/\varepsilon_{\mathrm{tr}}$ words, so that $M \gg 1$ words drawn from $[Q]_{\mathrm{tr}}$ will give rise to $\varepsilon_{\mathrm{tr}} M + o(M)$ such excursions. Moreover,

$$
[Q]_{\mathrm{tr}}\big(\varphi_{\mathrm{tr}}(Y^{(1,\dots,M)}) = \varphi_{\mathrm{tr}}(y)\big) \in \Big[\exp\big(-MH(\varphi_{\mathrm{tr}}([Q]_{\mathrm{tr}})) \pm \varepsilon_1 M\big)\Big], \tag{2.63}
$$

$$
[Q]_{\mathrm{tr}}\big(\widehat{\kappa}^*_{\mathrm{tr}}(Y^{(1,\dots,M)}) = \widehat{\kappa}^*_{\mathrm{tr}}(y)\big) \in \Big[\exp\big(-M\varepsilon_{\mathrm{tr}} H(\widehat{\kappa}^*_{\mathrm{tr}}([Q]_{\mathrm{tr}})) \pm \varepsilon_1 M\big)\Big]. \tag{2.64}
$$

(Note that both $\varphi_{\mathrm{tr}}([Q]_{\mathrm{tr}})$ and $\widehat{\kappa}^*_{\mathrm{tr}}(Q)$ are ergodic.)

Next, put

$$\mathscr{A}^* = \varphi_{\mathrm{tr}}(\mathscr{A}) \subset \widetilde{E}^M_{\mathrm{tr}}. \tag{2.65}$$

Consider the equivalence relation on $\mathscr{A} \subset \widetilde{E}^M$ defined by $y \sim y' \iff \varphi_{\mathrm{tr}}(y) = \varphi_{\mathrm{tr}}(y')$ (i.e., y and y' contain exactly the same short words at exactly the same positions, but contain possibly different long words), and identify $\mathscr{A}^*$ with a set of representatives (for example, we could replace each occurrence of the symbol $*_{\geq \mathrm{tr}}$ by a fixed string of tr letters). Below, we will sometimes implicitly use this convention when summing (or otherwise quantifying) over all $y \in \mathscr{A}^*$. Observe that $\widehat{\kappa}^*_{\mathrm{tr}}$ is well-defined on such equivalence classes.

Clearly, (2.63) and (2.64) imply that, for $y \in \mathscr{A}^*$,

$$\#\big\{y' \in \mathscr{A}^* : \widehat{\kappa}^*_{\mathrm{tr}}(y') = \widehat{\kappa}^*_{\mathrm{tr}}(y)\big\} \leq \exp\Big(M\big(H(\varphi_{\mathrm{tr}}([Q]_{\mathrm{tr}})) - \varepsilon_{\mathrm{tr}} H(\widehat{\kappa}^*_{\mathrm{tr}}([Q]_{\mathrm{tr}})) + 2\varepsilon_1\big)\Big), \tag{2.66}$$

because

$$\sum_{\substack{y' \in \mathscr{A}^*: \\ \widehat{\kappa}^*_{\mathrm{tr}}(y') = \widehat{\kappa}^*_{\mathrm{tr}}(y)}} [Q]_{\mathrm{tr}}\big(\varphi_{\mathrm{tr}}(Y^{(1,\dots,M)}) = y'\big) \leq [Q]_{\mathrm{tr}}\big(\widehat{\kappa}^*_{\mathrm{tr}}(Y^{(1,\dots,M)}) = \widehat{\kappa}^*_{\mathrm{tr}}(y)\big). \tag{2.67}$$

Combined with (2.41) from Lemma 2.12, (2.66) implies that

$$\#\big\{y' \in \mathscr{A}^* : \widehat{\kappa}^*_{\mathrm{tr}}(y') = \widehat{\kappa}^*_{\mathrm{tr}}(y)\big\} \leq \exp\Big(M\big(\varepsilon_{\mathrm{tr}} H_{\widehat{\tau}|\widehat{\kappa}^*_{\mathrm{tr}}}([Q]_{\mathrm{tr}}) + 2\varepsilon_1\big)\Big) \qquad \forall\, y \in \mathscr{A}^*. \tag{2.68}$$

Next, let

$$\mathscr{B} := \{\widehat{\kappa}^*_{\mathrm{tr}}(y) \colon\ y \in \mathscr{A}^*\}. \tag{2.69}$$

Then

$$\#\mathscr{B} \in \Big[\exp\big(M\varepsilon_{\mathrm{tr}} H\big(\widehat{\kappa}^*_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) \pm \varepsilon_1 M\big)\Big] \tag{2.70}$$

by ergodicity of $\widehat{\kappa}^*_{\mathrm{tr}}([Q]_{\mathrm{tr}})$, and, for $z \in \mathscr{B}$,

$$\begin{aligned} |z| &\in \big[\varepsilon_{\mathrm{tr}} M \ \pm\ \varepsilon_1 \varepsilon_{\mathrm{tr}} M\big], \\ \sum_{i=1}^{|\kappa(z)|} \log \nu(\kappa(z)_i) &\in \Big[M(1-\varepsilon_{\mathrm{tr}})\mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[\textstyle\sum_{i=1}^{|Y^{(1)}|} \log \nu(Y^{(1)}_i) \,\Big|\, |Y^{(1)}| < \mathrm{tr}\Big] \ \pm\ \varepsilon_1 M\Big] \end{aligned} \tag{2.71}$$

by ergodicity of $\varphi_{\mathrm{tr}}(Q)$. Properties (2.70)–(2.71) imply that

$$\begin{aligned} &\sum_{(\widetilde{z}_1,\dots,\widetilde{z}_\ell) \in \mathscr{B}} \prod_{j=1}^{\ell} \mathbb{P}(X \text{ begins with } \widetilde{z}_j) \\ &\qquad = \sum_{(\widetilde{z}_1,\dots,\widetilde{z}_\ell) \in \mathscr{B}} \prod_{j=1}^{\ell} \prod_{k=1}^{|\widetilde{z}_j|} \nu\big((\widetilde{z}_j)_k\big) \in \Big[\exp\Big(-Mm_{[Q]_{\mathrm{tr}}} H\big(\Psi_{[Q]_{\mathrm{tr}}} \,\big|\, \nu^{\otimes\mathbb{N}}\big) \pm 3\varepsilon_1 M\Big)\Big] \end{aligned} \tag{2.72}$$

when tr is sufficiently large, where we use that

$$\varepsilon_{\mathrm{tr}} H\big(\widehat{\kappa}^*_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big) + \mathbb{E}_{[Q]_{\mathrm{tr}}}\Big[\sum_{i=1}^{|Y^{(1)}|} \log \nu(Y_i^{(1)}) \,\Big|\, |Y^{(1)}| < \mathrm{tr}\Big] = -m_{[Q]_{\mathrm{tr}}} H\big(\Psi_{[Q]_{\mathrm{tr}}} \,\big|\, \nu^{\otimes\mathbb{N}}\big) + o(1) \tag{2.73}$$

as $\mathrm{tr} \to \infty$ by (2.52) from Proposition 2.15.

Finally, for $z \in \mathscr{B} = \widehat{\kappa}^*_{\mathrm{tr}}(\mathscr{A})$, set

$$f^*_{\mathrm{tr}}(z) = \sum_{\substack{y=(y^{(1)},\dots,y^{(M)})\in\mathscr{A}^*:\\ \widehat{\kappa}^*_{\mathrm{tr}}(y)=z}} \;\prod_{\substack{i=1,\dots,M:\\ y^{(i)}\in\widetilde{E}}} \rho(|y^{(i)}|). \tag{2.74}$$

We see from (2.68) and (2.62), together with (2.53) from Proposition 2.15, that

$$\log\big(f^*_{\mathrm{tr}}(z)\big) \le M\Big(H_{\tau|K}([Q]_{\mathrm{tr}}) + (1-\varepsilon_{\mathrm{tr}})\mathbb{E}_{[Q]_{\mathrm{tr}}}\big[\log\rho(\tau_1)\,\big|\,\tau_1 < \mathrm{tr}\big] + 4\varepsilon_1\Big) \tag{2.75}$$

for all $z \in \mathscr{B}$.

2.5 Empirical process

Define

$$\mathcal{O} = \big\{Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) : (\pi_M Q')\big(\varphi_{\mathrm{tr}}^{-1}(\mathscr{A}^*)\big) > 1-\delta_1\big\} \tag{2.76}$$

with $\delta_1 > 0$ as in (2.60) (later, we will let $\delta_1 \downarrow 0$). For $N \in \mathbb{N}$, $1 \le j_1 < j_2 < \dots < j_N$ and a (possibly random) sequence X of letters from E, let $\xi_N = \xi_N(X; j_1, \dots, j_N)$ be defined as in [BGdH10, Eq. (3.16)], and let $R^N_{j_1,\dots,j_N}(X)$ be defined as in [BGdH10, Eq. (3.17)]:

$$\xi_N := (\xi^{(i)})_{i=1,\dots,N} = \big(X|_{(0,j_1]}, X|_{(j_1,j_2]}, \dots, X|_{(j_{N-1},j_N]}\big) \in \widetilde{E}^N, \tag{2.77}$$

$$R^N_{j_1,\dots,j_N}(X) := \frac{1}{N}\sum_{i=0}^{N-1} \delta_{\widetilde{\theta}^i(\xi_N)^{\mathrm{per}}}. \tag{2.78}$$

This is the empirical process of the sequence of words obtained by cutting X at the positions j_i, where $(0, j_1]$ is a shorthand notation for $(0, j_1] \cap \mathbb{N}$, etc. As in [BGdH10, Section 3.2, Eq. (3.19)], $R^N_{j_1,\dots,j_N}(x) \in \mathcal{O}$ implies that there are at least

$$\widetilde{N} = \lfloor (1-\delta_1)N/M \rfloor - 1 \tag{2.79}$$

positions $1 \le k_1 < \dots < k_{\widetilde{N}}$ with $k_i - k_{i-1} \ge M$ such that

$$\varphi_t\big((\xi^{(k_i)}, \xi^{(k_i+1)}, \dots, \xi^{(k_i+M-1)})\big) \in \mathscr{A}^*, \qquad i = 1, 2, \dots, \widetilde{N}. \tag{2.80}$$

Then, by (2.61),

$$\widehat{\kappa}^*_{\mathrm{tr}}\big((\xi^{(k_i)}, \xi^{(k_i+1)}, \dots, \xi^{(k_i+M-1)})\big) =: z_i = \big(\widetilde{z}_{i,1}, \dots, \widetilde{z}_{i,\ell_i}\big) \in (E^*)^{\ell_i} \tag{2.81}$$

with $\ell_i = |z_i| \in [\varepsilon_{\mathrm{tr}} M \pm \varepsilon_1 \varepsilon_{\mathrm{tr}} M]$, i.e., z_i consists of ℓ_i excursions from $E^{\ge \mathrm{tr}}$ and we denote by $\widetilde{z}_{i,j} \in E^*$ the concatenation of the short words inside the j-th excursion contained in z_i. Note that, by construction, $z_i \in \mathscr{B}$ with $\mathscr{B}$ from (2.69).

For the mapping from $(j_1,\dots,j_N)$ with $R^N_{j_1,\dots,j_N}(X)\in\mathcal{O}$ to positions on the sequence X (where certain patterns must occur), we use the following notation. Write $p_{i,1}<\dots<p_{i,\ell_i}$ for the positions (in X) where the excursions $\widetilde{z}_{i,1},\dots,\widetilde{z}_{i,\ell_i}$ start (i.e., $\widetilde{z}_{i,r}$ is a prefix of $\theta^{p_{i,r}}(X)$). Note that necessarily

$$p_{i,r+1}\ge p_{i,r}+|\widetilde{z}_{i,r}|+\mathrm{tr},\quad r<\ell_i,\, i\le\widetilde{N},\qquad p_{i+1,1}\ge p_{i,\ell_i}+|\widetilde{z}_{i,\ell_i}|,\quad i<\widetilde{N},\tag{2.82}$$

and $\bigcup_{i=1}^{\widetilde{N}}\{p_{i,1},\dots,p_{i,\ell_i}\}\subset\{j_1,\dots,j_N\}$. Put

$$\begin{aligned}S_N:=\sum_{(z_1,\dots,z_{\widetilde{N}})\in\mathscr{B}^{\widetilde{N}}}\ \sum_{\substack{(p_{i,r})\\ \text{compatible}\\ \text{with }(z_i)}}\ \sum_{\substack{(s_1,\dots,s_{\widetilde{N}})\in\mathbb{N}_0^{\widetilde{N}}\\ s_1+\dots+s_{\widetilde{N}}\le N-M\widetilde{N}}}\prod_{i=1}^{\widetilde{N}}\Big(&\rho^{*s_i}\big(p_{i,1}-p_{i-1,\ell_{i-1}}-|\widetilde{z}_{i-1,\ell_{i-1}}|\big)f^*_{\mathrm{tr}}(z_i)\\ &\times\prod_{r=1}^{\ell_i-1}\Big(\mathbb{1}\big(\theta^{p_{i,r}}(X)\text{ begins with }\widetilde{z}_{i,r}\big)\rho\big(p_{i,r+1}-p_{i,r}-|\widetilde{z}_{i,r}|\big)\Big)\Big)\\ &\times\mathbb{1}\big(\theta^{p_{i,\ell_i}}(X)\text{ begins with }\widetilde{z}_{i,\ell_i}\big)\end{aligned}\tag{2.83}$$

In the first sum, $z_i=(\widetilde{z}_{i,1},\dots,\widetilde{z}_{i,\ell_i})\in\mathscr{B}$ is composed of ℓ_i "stretches". Some $\widetilde{z}_{i,r}$ may be equal to the empty word ϵ, in which case we put $\mathbb{1}\big(\theta^{p_{i,r}}(X)\text{ begins with }\widetilde{z}_{i,r}\big)\equiv 1$. In the second sum, $(p_{i,r})$ is compatible with a given choice of $(z_1,\dots,z_{\widetilde{N}})$ when it respects (2.82). In the third sum, s_i describes how many words the empirical process cuts out between the occurrence of z_{i-1} and z_i. The condition $R^N_{j_1,\dots,j_N}(X)\in\mathcal{O}$ enforces that $s_1+\dots+s_{\widetilde{N}}\le N-M\widetilde{N}\le\delta_1 N$. Recalling from [BGdH10, Eq. (3.23)] that $\mathbb{P}(R_N\in\mathcal{O}\mid X)$ is a weighted sum of loop configurations, we have

$$\mathbb{P}(R_N\in\mathcal{O}\mid X)\le S_N.\tag{2.84}$$

Let us for the moment assume that ρ satisfies the bound

$$\rho(n)\le C_\rho n^{-\alpha},\quad n\in\mathbb{N},\tag{2.85}$$

which is stronger than (1.1). Then, by [BGdH10, Lemma 2.3], we have $\rho^{*s}(n)\le(C_\rho\vee 1)s^{\alpha+1}n^{-\alpha}$, and so

$$\sum_{\substack{(s_1,\dots,s_{\widetilde{N}})\in\mathbb{N}_0^{\widetilde{N}}\\ s_1+\dots+s_{\widetilde{N}}\le N-M\widetilde{N}}}\prod_{i=1}^{\widetilde{N}}\rho^{*s_i}\big(p_{i,1}-p_{i-1,\ell_{i-1}}-|\widetilde{z}_{i-1,\ell_{i-1}}|\big)\le e^{\varepsilon(N)N}\prod_{i=1}^{\widetilde{N}}\big(p_{i,1}-p_{i-1,\ell_{i-1}}-|\widetilde{z}_{i-1,\ell_{i-1}}|\big)^{-\alpha},\tag{2.86}$$

where we use that

$$\begin{aligned}(C_\rho\vee 1)^{\widetilde{N}}\sum_{\substack{(s_1,\dots,s_{\widetilde{N}})\in\mathbb{N}_0^{\widetilde{N}}\\ s_1+\dots+s_{\widetilde{N}}\le N-M\widetilde{N}}}\prod_{i=1}^{\widetilde{N}}s_i^{\alpha+1}&\le(C_\rho\vee 1)^{\widetilde{N}}\Big(\frac{\delta_1 N}{\widetilde{N}}\Big)^{\widetilde{N}(\alpha+1)}\sum_{u=\widetilde{N}}^{N-M\widetilde{N}}\binom{N-M\widetilde{N}+u-1}{u-1}\\ &=:\mathrm{e}^{\varepsilon(N)N}.\end{aligned}\tag{2.87}$$

Note that $\limsup_{N\to\infty}\varepsilon(N)$ can be made arbitrarily small by letting $M\to\infty$ and $\delta_1\downarrow 0$.

2.6 Estimating the cost of finding suitable stretches in the given medium

A fractional moment estimate similar to the one used in the proof of [BGdH10, Lemma 2.1] suffices to conclude the proof of Theorem 2.3. Pick $1/\alpha < b < 1$. Using $(\sum_i x_i)^b \le \sum_i x_i^b$ for $x_i \ge 0$, we estimate the b-th moment of S_N from (2.83) as follows:

$$
\begin{aligned}
\mathbb{E}[S_N^b] \le\; & \mathrm{e}^{\varepsilon(N)N} \sum_{(z_1,\dots,z_{\widetilde{N}})\in\mathscr{B}^{\widetilde{N}}} \ \sum_{\substack{(p_{i,r}) \\ \text{compatible with } (z_i)}} \ \prod_{i=1}^{\widetilde{N}} \Bigg(\big(p_{i,1} - p_{i-1,\ell_{i-1}} - |\widetilde{z}_{i-1,\ell_{i-1}}|\big)^{-b\alpha} \big(f_{\mathrm{tr}}^*(z_i)\big)^b \\
& \times \prod_{r=1}^{\ell_i-1} \Big(\mathbb{P}(X \text{ begins with } \widetilde{z}_{i,r}) \big(p_{i,r+1} - p_{i,r} - |\widetilde{z}_{i,r}|\big)^{-b\alpha} \Big) \Bigg) \times \mathbb{P}(X \text{ begins with } \widetilde{z}_{i,\ell_i}) \\
\le\; & \mathrm{e}^{\varepsilon(N)N} \exp\bigg(\widetilde{N} M b \Big(H_{\tau|K}([Q]_{\mathrm{tr}}) + (1-\varepsilon_{\mathrm{tr}}) \mathbb{E}_{[Q]_{\mathrm{tr}}}\big[\log\rho(\tau_1) \,\big|\, \tau_1 < \mathrm{tr}\big] + 4\varepsilon_1 \Big) \bigg) \\
& \times \sum_{z_1,\dots,z_{\widetilde{N}}\in\mathscr{B}} \Big(\sum_{p=1}^{\infty} p^{-b\alpha} \Big)^{\widetilde{N}} \Big(\sum_{p=\mathrm{tr}}^{\infty} p^{-b\alpha} \Big)^{\sum_{i=1}^{\widetilde{N}} \#\{j \le \ell_i : \widetilde{z}_{i,j} = \epsilon\}} \times \prod_{i=1}^{\widetilde{N}} \prod_{r=1}^{\ell_i} \mathbb{P}(X \text{ begins with } \widetilde{z}_{i,r}) \\
\le\; & \mathrm{e}^{\varepsilon(N)N} \big(\zeta(b\alpha)\big)^{\widetilde{N}} \exp\bigg(\widetilde{N} M b \Big(H_{\tau|K}([Q]_{\mathrm{tr}}) + (1-\varepsilon_{\mathrm{tr}}) \mathbb{E}_{[Q]_{\mathrm{tr}}}\big[\log\rho(\tau_1) \,\big|\, \tau_1 < \mathrm{tr}\big] + 5\varepsilon_1 \Big) \\
& \qquad - \widetilde{N} M m_{[Q]_{\mathrm{tr}}} H\big(\Psi_{[Q]_{\mathrm{tr}}} \,\big|\, \nu^{\otimes\mathbb{N}}\big) \bigg) \\
\le\; & \exp\bigg(N \Big(b\big(H_{\tau|K}([Q]_{\mathrm{tr}}) + \mathbb{E}_{[Q]_{\mathrm{tr}}}\big[\log\rho(\tau_1)\big]\big) - m_{[Q]_{\mathrm{tr}}} H(\Psi_{[Q]_{\mathrm{tr}}} \mid \nu^{\otimes\mathbb{N}}) + \varepsilon_2 \Big) \bigg),
\end{aligned}
\tag{2.88}
$$

where ε_2 can be made arbitrarily small by taking $\mathrm{tr} \to \infty$. Here, we use (2.75) in the first inequality and (2.72) in the third inequality. Thus, with

$$
c := 2\varepsilon_2/b + \big(\mathbb{E}_{[Q]_{\mathrm{tr}}}\big[\log\rho(\tau_1)\big] + H_{\tau|K}([Q]_{\mathrm{tr}}) \big) + \frac{m_{[Q]_{\mathrm{tr}}}}{b} H(\Psi_{[Q]_{\mathrm{tr}}} \mid \nu^{\otimes\mathbb{N}}), \tag{2.89}
$$

we have

$$
\mathbb{P}\Big(\frac{1}{N} \log S_N \ge c \Big) = \mathbb{P}\Big(S_N^b \ge \mathrm{e}^{bcN} \Big) \le \mathrm{e}^{-bcN} \mathbb{E}\big[S_N^b\big] \le \mathrm{e}^{-\varepsilon_2 N}, \tag{2.90}
$$

and hence, by Borel-Cantelli,

$$
\begin{aligned}
& \limsup_{N\to\infty} \frac{1}{N} \log S_N \\
& \quad \le 2\varepsilon_2/b + \big(\mathbb{E}_{[Q]_{\mathrm{tr}}}\big[\log\rho(\tau_1)\big] + H_{\tau|K}([Q]_{\mathrm{tr}}) \big) - \frac{m_{[Q]_{\mathrm{tr}}}}{b} H(\Psi_{[Q]_{\mathrm{tr}}} \mid \nu^{\otimes\mathbb{N}}) \quad \text{a.s.}
\end{aligned}
\tag{2.91}
$$

Taking $1 \gg b - (1/\alpha) > 0$ and recalling (2.6), we get

$$
\limsup_{N\to\infty} \frac{1}{N} \log S_N \le \delta - I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) \quad \text{a.s.}, \tag{2.92}
$$

where δ can be chosen arbitrarily small by taking tr sufficiently large. Combining (2.84) and (2.92), we complete the proof of Theorem 2.3.

Finally, if ρ does not satisfy (2.85) but only (1.1), then we can estimate $\rho(n) \le C(\alpha')n^{-\alpha'}$ with $1 < \alpha' < \alpha$, argue as above and let $\alpha' \uparrow \alpha$, analogous to the argument in [BGdH10, Section 3.6].

With Sections 2.3–2.6, we have completed the proof of Theorem 1.5 for finite E and ergodic Q.

2.7 Extension to Polish alphabets for ergodic measures

Let the letter space E be Polish. The arguments given in [BGdH10, Section 8] are correct for Q with $m_Q < \infty$, i.e., up to and including [BGdH10, Eq. (8.14)]. In particular, the set-up with a nested sequence of colour-coarse-grainings $\langle\cdot\rangle_c \colon E \to \mathscr{A}_c \,(= \langle E\rangle_c)$, $c \in \mathbb{N}$, and the corresponding projective limit F as defined in [BGdH10, Eq. (8.2)] are sound. But [BGdH10, Eq. (8.15)] considers $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{F}^{\mathbb{N}})$ with $m_Q = \infty$ and is not correct as it stands. In fact, it employs the erroneous claim that $I^{\mathrm{que}}(Q) = \sup_{\mathrm{tr}\in\mathbb{N}} I^{\mathrm{fin}}([Q]_{\mathrm{tr}})$. The arguments after [BGdH10, Eq. (8.15)] at the end of [BGdH10, Section 8], where the LDP is transferred from $\mathcal{P}^{\mathrm{inv}}(\widetilde{F}^{\mathbb{N}})$ to $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$, are again correct. This shows that Theorem 1.3 does indeed hold for a general Polish state space E, and so we have to show that (1.24) also holds for a general Polish state space E.

For the moment, consider $Q \in \mathcal{P}^{\mathrm{erg}}(\widetilde{F}^{\mathbb{N}})$ with $m_Q = \infty$. We show that

$$\widetilde{I}^{\mathrm{que}}(Q) \geq \limsup_{\mathrm{tr}\to\infty} I^{\mathrm{fin}}([Q]_{\mathrm{tr}}). \tag{2.93}$$

The proof of (2.93) is achieved by a careful re-reading of our arguments in Sections 2.3–2.6, as follows. For a given $c \in \mathbb{N}$, apply the constructions in Section 2.3 to $\langle Q\rangle_c$, which has the finite letter space $\langle E\rangle_c = \langle F\rangle_c$. We extend $\langle\cdot\rangle_c$ to act on words, word sequences, measures, etc. as in [BGdH10, Section 8], and we stipulate that $\langle *_{\geq \mathrm{tr}}\rangle_c = *_{\geq \mathrm{tr}}$ for any c. Since φ_{tr} and $\langle\cdot\rangle_c$ are consistent families of commuting projections, we have

$$H\big(Q \,\big|\, Q_{\rho,\nu}\big) = \sup_{c\in\mathbb{N},\,\mathrm{tr}\in\mathbb{N}} H\big(\varphi_{\mathrm{tr}}(\langle Q\rangle_c) \,\big|\, \varphi_{\mathrm{tr}}(\langle Q_{\rho,\nu}\rangle_c)\big). \tag{2.94}$$

This shows that the error terms $o(1)$ in (2.36) and (2.37) in Corollary 2.11 can be taken small jointly uniformly in c and tr, in the sense that

$$\forall\,\delta > 0 :\ \exists\, c_0 \in \mathbb{N}, \mathrm{tr}_0 \in \mathbb{N} :\ \forall\, \mathrm{tr} \geq \mathrm{tr}_0, c \geq c_0 :\ \big|o(1)\big| \leq \delta. \tag{2.95}$$

We obtain from this that the error terms $o(1/\varepsilon_{\mathrm{tr}})$ in (2.43) from Lemma 2.13 and in (2.49) from Lemma 2.14 can be quantified jointly in c and tr, in the sense that

$$\forall\,\delta > 0 :\ \exists\, c_0 \in \mathbb{N}, \mathrm{tr}_0 \in \mathbb{N} :\ \forall\, \mathrm{tr} \geq \mathrm{tr}_0, c \geq c_0 :\ \big|o(1/\varepsilon_{\mathrm{tr}})\big| \leq \frac{\delta}{\varepsilon_{\mathrm{tr}}}. \tag{2.96}$$

Indeed, in the arguments in Section 2.3 no properties of the letter space are required (apart from being finite). In particular, no bound derived there depends on the size of the letter space.

Feeding (2.95) and (2.96) into the proof of Proposition 2.15, we see that the error terms $o(1)$ in both (2.52) and (2.53) from Proposition 2.15 satisfy the joint quantification described in (2.95).

This shows that for given $\delta > 0$ we can can pick large enough c_0 and tr_0 and carry through the arguments from Sections 2.4–2.6 for $\langle Q\rangle_{c_0}$ with the finite letter space $\langle E\rangle_{c_0}$, in order to construct for any $\mathrm{tr} \geq \mathrm{tr}_0$ an open neighbourhood $\mathcal{O} \subset \mathcal{P}^{\mathrm{inv}}(\widetilde{F}^{\mathbb{N}})$ of Q as in (2.76) with the property that

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in \mathcal{O} \mid X) \leq -I^{\mathrm{fin}}_{c_0}([\langle Q\rangle_{c_0}]_{\mathrm{tr}}) + \delta \leq -I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) + 2\delta, \tag{2.97}$$

where

$$I^{\mathrm{fin}}_c(Q) = H(Q \mid Q_{\rho,\langle\nu\rangle_c}) + m_Q H(\Psi_Q \mid \langle\nu\rangle_c^{\otimes\mathbb{N}}) \tag{2.98}$$

for $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{\langle E\rangle_c})$ (note that $I^{\mathrm{fin}}_c([\langle Q\rangle_c]_{\mathrm{tr}}) \uparrow I^{\mathrm{fin}}([Q]_{\mathrm{tr}})$ as $c \to \infty$). This implies (2.93).

2.8 Extension to non-ergodic measures

In the following, let the letter space E be a Polish space and fix a nested sequence of colour-coarse-grainings $\langle\cdot\rangle_c\colon E \to \langle E\rangle_c$, $c \in \mathbb{N}$, into finitely many colours as in Section 2.7. Let $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ have a non-trivial ergodic decomposition

$$Q = \int_{\mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})} Q'\, W_Q(dQ'), \tag{2.99}$$

where W_Q is a probability measure on $\mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})$ that is not concentrated on a single point (see e.g. Georgii [Ge88, Proposition 7.22]). In order to construct a neighbourhood $\mathcal{O} \ni Q$ satisfying (2.10), we essentially re-read the approach from [BGdH10, Section 3.5] in the light of the ideas and constructions from Sections 2.3–2.7 above. Very broadly, we approximate the integral in (2.99) by a finite convex combination

$$Q \approx \sum_{r=1}^{R} \lambda_r Q_r \tag{2.100}$$

with pairwise disjoint $Q_r \in \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})$. We choose $Q_1, \dots, Q_R$ so that, in particular,

$$I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) \approx \sum_{r=1}^{R} \lambda_r I^{\mathrm{fin}}\big([Q_r]_{\mathrm{tr}}\big) \tag{2.101}$$

is approximated in a suitably quantified way. It will be important that (2.100)–(2.101) can be made precise uniformly in (sufficiently) large colour-coarse-graining levels c. Sections 2.8.1–2.8.3 detail the steps in the construction of $\mathcal{O} \ni Q$.

2.8.1 Preliminary observations and tools

We may assume w.l.o.g. that $I^{\mathrm{ann}}(Q) = H(Q \mid Q_{\rho,\nu}) < \infty$, otherwise we can simply employ the annealed bound. Thus,

$$W_Q\big(\mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}}) \cap \{Q'\colon\ H(Q' \mid Q_{\rho,\nu}) < \infty\}\big) = 1. \tag{2.102}$$

Remark 2.16. [Upper bound on the rate function] Observe that

$$\sup_{\mathrm{tr}\in\mathbb{N}} I^{\mathrm{fin}}([Q']_{\mathrm{tr}}) \le \alpha\, H(Q' \mid Q_{\rho,\nu}) \qquad \forall\, Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}). \tag{2.103}$$

(Note that (2.103) does not require any condition on the underlying space E.) Indeed, applying (2.7) with ρ replaced by $[\rho]_{\mathrm{tr}}$, we see that $H([Q']_{\mathrm{tr}} \mid [Q_{\rho,\nu}]_{\mathrm{tr}}) \ge m_{[Q']_{\mathrm{tr}}} H(\Psi_{[Q']_{\mathrm{tr}}} \mid \nu^{\otimes\mathbb{N}})$, because the difference term $-\mathbb{E}_{[Q]_{\mathrm{tr}}}[\log([\rho]_{\mathrm{tr}}(\tau_1))] - H_{\tau|K}([Q]_{\mathrm{tr}}) \ge 0$ is a conditional specific relative entropy. Combined with (2.6), this yields $I^{\mathrm{fin}}([Q']_{\mathrm{tr}}) \le \alpha H([Q']_{\mathrm{tr}} \mid [Q_{\rho,\nu}]_{\mathrm{tr}}) \le \alpha H(Q' \mid Q_{\rho,\nu})$. $\triangle$

We summarize relevant constructions and results from Sections 2.5, 2.6 and 2.7 in the following lemma.

Lemma 2.17. [Approximation by open neighbourhoods] *Let $Q \in \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})$ and $\delta > 0$. There exist $\mathrm{tr}_0 = \mathrm{tr}_0(Q,\delta)$, $c_0 = c_0(Q,\delta)$ and $M_0 = M_0(Q,\delta) \in \mathbb{N}$ with the following properties: For $M \ge M_0$, $c \ge c_0$ and $\mathrm{tr} \ge \mathrm{tr}_0$ it is possible to construct an open neighbourhood $\mathcal{U}_Q \subset \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$*

of Q based on an M-cylinder set $\mathscr{A}^* \subset \big(\widetilde{\langle E \rangle}_c \cup \{*_{\geq \mathrm{tr}}\}\big)^M$ of words that are colour-coarse-grained to c colours, and with words of lengths $\geq \mathrm{tr}$ replaced by a special symbol as in (2.76)*, for which the estimates* (2.70)*,* (2.71)*,* (2.72)*,* (2.75) *hold (after $[Q]_{\mathrm{tr}}$ is replaced by $[\langle Q \rangle_c]_{\mathrm{tr}}$ in the formulas). In fact, it is possible to take a neighbourhood of the form*

$$\mathcal{U}_Q = \Big\{ Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) \,:\, \frac{\pi_M \varphi_{\mathrm{tr}}(\langle Q' \rangle_c)(z)}{\pi_M \varphi_{\mathrm{tr}}(\langle Q \rangle_c)(z)} \in \big(1 \pm \delta_1\big) \text{ for all } z \in \mathscr{A}^* \Big\} \tag{2.104}$$

with a suitable $\delta_1 = \delta_1(Q) > 0$. In particular, the arguments from Sections 2.5 and 2.6 yield

$$\mathbb{P}\big(R_N \in \overline{\mathcal{U}_Q} \,\big|\, X\big) \leq \exp\big(-N(I_c^{\mathrm{fin}}([\langle Q \rangle_c]_{\mathrm{tr}}) - \delta)\big) \tag{2.105}$$

for all N sufficiently large, with I_c^{fin} as in (2.98)*.*

2.8.2 Constructing a suitable neighbourhood of Q

Fix $\delta > 0$, which will (up to a fixed numerical constant) play the role of δ on the right-hand side of (2.10). For $c, \mathrm{tr}, M \in \mathbb{N}$, put

$$\mathcal{P}(c, \mathrm{tr}, M) := \left\{ Q \in \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}}) \;:\; \begin{array}{l} \text{There exists a neighbourhood } \mathcal{U}_Q \ni Q \text{ as in (2.104) in} \\ \text{Lemma 2.17 with this choice of } c, \mathrm{tr}, M \text{ and a suitable} \\ \delta_1 > 0 \text{ satisfying } \inf_{Q' \in \mathcal{U}_Q} I_c^{\mathrm{fin}}([\langle Q' \rangle_c]_{\mathrm{tr}}) \geq I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) - \delta \end{array} \right\}. \tag{2.106}$$

Lemma 2.17 together with lower semi-continuity of I^{fin} gives $\bigcup_{(c,\mathrm{tr},M) \in \mathbb{N}^3} \mathcal{P}(c, \mathrm{tr}, M) = \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})$. Fix c, tr, M_1 so large that

$$\int_{\mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}}) \setminus \mathcal{P}(c, \mathrm{tr}, M_1)} I^{\mathrm{fin}}([Q']_{\mathrm{tr}}) \, W_Q(dQ') < \delta \tag{2.107}$$

(since $I^{\mathrm{ann}}(Q) < \infty$ this is possible in view of Remark 2.16). Using Remark 2.16 and $I^{\mathrm{ann}}(Q) < \infty$ again, we can choose $0 < K < \infty$ so large that

$$\int_{\{Q' \,:\, I^{\mathrm{fin}}(Q') > K\}} I^{\mathrm{fin}}([Q']_{\mathrm{tr}}) \, W_Q(dQ') < \delta.$$

Now $\overline{\mathcal{P}(c, \mathrm{tr}, M_1)} \cap \{Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) : I^{\mathrm{ann}}(Q') \leq K\}$ is a compact subset of $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ and

$$\bigcup_{Q' \in \mathcal{P}(c, \mathrm{tr}, M_1)} \mathcal{U}_{Q'} \supset \overline{\mathcal{P}(c, \mathrm{tr}, M_1)} \cap \{Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) : I^{\mathrm{ann}}(Q') \leq K\} \tag{2.108}$$

is an open cover.

Thus, we can find $R \in \mathbb{N}$, centers $Q_1, \dots, Q_R \in \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}}) \cap \mathrm{supp}(W_Q)$ and finite sets $\mathscr{A}_{1,r}^* \subset \langle \widetilde{E}_{\mathrm{tr}} \rangle_c^{M_1}$ of "Q_r-typical M_1-sentences" (more precisely: $\varphi_{\mathrm{tr}}(\langle Q_r \rangle_c)$-typical M_1-sentences) with the properties from Section 2.4 and use them to construct, for $r = 1, \dots, R$, a neighbourhood $\mathcal{U}_r$ of Q_r of the form (2.104) in Lemma 2.17 with a suitably tuned $0 < \delta_1 = \delta_1(Q_r) \ll 1$, analogous to [BGdH10, Eqs. (3.71) and Eq. (3.92)], such that

$$I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) \leq \int_{\cup_{r=1}^R \mathcal{U}_r} I^{\mathrm{fin}}([\langle Q' \rangle_c]_{\mathrm{tr}}) \, W_Q(dQ') + 3\delta. \tag{2.109}$$

Note that $\mathscr{B}_r := \widehat{\kappa}^*_{\mathrm{tr}}(\mathscr{A}^*_{1,r})$ fulfils (the analogues of) (2.72) and (2.75) for $\langle Q_r\rangle_c$, $r = 1,\dots,R$, with the same ε_1, which we fix here (we will take $\varepsilon_1 \downarrow 0$ later). It is important that we can choose M_1, c and tr such that this works simultaneously for all Q_r.

We 'disjointify' $\mathcal{U}_r$, $r = 1,\dots,R$, analogously to [BGdH10, Eq. (3.98)] by defining, iteratively,

$$\widetilde{\mathcal{U}}_r = \left\{ Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) \,:\, \begin{array}{l} \pi_{M_1}\varphi_{\mathrm{tr}}(\langle Q'\rangle_c)(z)/\pi_{M_1}\varphi_{\mathrm{tr}}(\langle Q_r\rangle_c)(z) \in \big(1 \pm \delta_2(Q_r)\big) \text{ for all } z \in \mathscr{A}^*_{1,r} \\ \text{and for each } r' < r \text{ there exists } z' \in \mathscr{A}^*_{1,r'} \text{ such that} \\ \pi_{M_1}\varphi_{\mathrm{tr}}(\langle Q'\rangle_c)(z')/\pi_{M_1}\varphi_{\mathrm{tr}}(\langle Q_r\rangle_c)(z') \notin \big[1 \pm (1+\widetilde{\delta})\delta_2(Q_r)\big] \end{array} \right\} \tag{2.110}$$

with a carefully tuned $0 < \widetilde{\delta} \ll 1$ and a carefully tuned $\delta_2(Q_r)$ derived from $\delta_1(Q_r)$ as in [BGdH10, Section 3.5.3]. The point here is that we need to avoid possible atoms of the measures $W_Q(\{Q' : \pi_{M_1}\varphi_{\mathrm{tr}}(\langle Q\rangle_c)(z) \in \cdot\})$, $z \in \cup_{r=1}^R \mathscr{A}^*_{1,r}$, sitting exactly at the boundaries of the sets of inequalities implicit in (2.110) (see the discussion around [BGdH10, Eq. (3.96)]). We silently remove possibly resulting empty sets and re-number the remaining ones without making this notationally explicit. We can arrange the choices of $\delta_2(Q_r)$ and $\widetilde{\delta}$ in (2.110) in such a way that

$$I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) \le 4\delta + \sum_{r=1}^R \lambda_r I^{\mathrm{fin}}_c([\langle Q_r\rangle_c]_{\mathrm{tr}}) \tag{2.111}$$

with $\lambda_r := W_Q(\widetilde{\mathcal{U}}_r)$. This is a quantitative version of (2.101), analogous to [BGdH10, Eq. (3.102)].

Based on the M_1-pattern frequencies under Q_r from the definition of $\widetilde{\mathcal{U}}_r$, we choose $M_2 \gg M_1$ and define sets $\mathscr{A}^*_{2,r} \subset \langle \widetilde{E}_{\mathrm{tr}}\rangle_c^{M_2}$ of "Q_r-typical M_2-sentences" that have (approximately) the correct frequencies of all M_1-patterns from $\mathscr{A}^*_{1,r}$, parallel to [BGdH10, Section 3.5.4]: For $\xi = (\xi^{(1)}\dots,\xi^{(M_2)}) \in \langle \widetilde{E}_{\mathrm{tr}}\rangle_c^{M_2}$ and $z \in \langle \widetilde{E}_{\mathrm{tr}}\rangle_c^{M_1}$, write

$$\mathrm{freq}_z(\xi) := \frac{1}{M_2 - M_1 + 1}\#\big\{1 \le i \le M_2 - M_1 + 1 : (\xi^{(i)}\dots,\xi^{(i+M_1-1)}) = z\big\} \tag{2.112}$$

for the empirical frequency of z as a subword of ξ. Put

$$\mathscr{A}^*_{2,r} = \left\{ \xi \in \langle \widetilde{E}_{\mathrm{tr}}\rangle_c^{M_2} \,:\, \begin{array}{l} \mathrm{freq}_z(\xi)/\varphi_{\mathrm{tr}}(\langle Q_r\rangle_c)(z) \in \big(1 \pm \delta_2(Q_r)\big) \text{for all } z \in \mathscr{A}^*_{1,r} \\ \text{and for each } r' < r \text{ there exists } z' \in \mathscr{A}^*_{1,r'} \text{ such that} \\ \mathrm{freq}_{z'}(\xi)/\varphi_{\mathrm{tr}}(\langle Q_r\rangle_c)(z') \notin \big[1 \pm (1+\widetilde{\delta})\delta_2(Q_r)\big] \end{array} \right\} \tag{2.113}$$

with $\delta_2(Q_r), \widetilde{\delta}$ as in (2.110). This is the analogue of [BGdH10, Eq. (3.106)]. We choose $M_2 \gg M_1$ so large and $\varepsilon_2, \varepsilon_3 > 0$ suitably small ($\varepsilon_2, \varepsilon_3 \downarrow 0$ as $\delta \downarrow 0$ eventually) such that

$$W_Q\big(\{Q' \in \widetilde{\mathcal{U}}_r \cap \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}}) \colon \pi_{M_2}\varphi_{\mathrm{tr}}\big(\langle Q'\rangle_c\big)(\mathscr{A}^*_{2,r}) > 1 - \varepsilon_3\}\big) > W_Q(\widetilde{\mathcal{U}}_r)(1 - \varepsilon_2), \quad r = 1,\dots,R. \tag{2.114}$$

(cf. [BGdH10, Eq. (3.114)].)

Note that if, for some $r \neq r'$, Q_r and $Q_{r'}$ have many similar M_1-pattern frequencies, then it could happen that $\xi \in \mathscr{A}^*_{2,r}$ and $\xi' \in \mathscr{A}^*_{2,r'}$ have substantial overlap though not being identical. This would be a problem in the construction. To remedy this, we introduce another layer: For $M_3 \gg M_2$ (suitably tuned), define

$$\mathscr{A}^*_{3,r} = \big\{\zeta \in \langle \widetilde{E}_{\mathrm{tr}}\rangle_c^{M_3} : \mathrm{freq}_{\mathscr{A}^*_{2,r}}(\zeta) > 1 - 2\varepsilon_2\big\}, \quad r = 1,\dots,R, \tag{2.115}$$

where $\mathrm{freq}_{\mathscr{A}^*_{2,r}}(\zeta)$ is the relative frequency of positions in ζ at which a subword from $\mathscr{A}^*_{2,r}$ begins, analogous to [BGdH10, Eq. (3.115)]. (See the discussion in [BGdH10] around Eq. (3.116) to understand how this construction prevents "large overlaps".) We can make M_3 so large that

$$W_Q\big(\{Q' \in \widetilde{\mathscr{U}}_r \cap \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}}) \colon\ \pi_{M_3}\varphi_{\mathrm{tr}}\big(\langle Q'\rangle_c\big)(\mathscr{A}^*_{3,r}) > 1-2\varepsilon_2\}\big) > W_Q(\widetilde{\mathscr{U}}_r)(1-4\varepsilon_2), \qquad r=1,\dots,R. \tag{2.116}$$

(See [BGdH10, Eq. (3.120)] and the preceding discussion) and then define

$$\mathcal{O} := \big\{Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}) : \pi_{M_3}\varphi_{\mathrm{tr}}\big(\langle Q'\rangle_c\big)(\mathscr{A}^*_{3,r}) > W_Q(\widetilde{\mathscr{U}}_r)(1-5\varepsilon_2) \text{ for } r=1,\dots,R\big\} \tag{2.117}$$

analogous to [BGdH10, Eq. (3.121)]. This $\mathcal{O} \ni Q$ will be the open neighbourhood we work with to prove (2.10).

For N sufficiently large and cut-points $1 \le j_1 < j_2 < \cdots < j_N$ as in (2.77)–(2.78), $R^N_{j_1,\dots,j_N}(X) \in \mathcal{O}$ from (2.117) implies that $\varphi_{\mathrm{tr}}(\langle \xi_N\rangle_c)$ contains at least

$$\widetilde{N} := \lfloor (1-\varepsilon)N/M_1 \rfloor \tag{2.118}$$

disjoint occurrences of subsentences from $\cup_{r=1}^R \mathscr{A}^*_{1,r}$ with "approximately correct frequencies", i.e., there are $0 \le i_1 < \cdots < i_{\widetilde{N}} \le N - M_1$ with $i_{k+1} \ge i_k + M_1$ such that (analogous to [BGdH10, Eq. (3.131)])

$$\#\big\{1 \le k \le \widetilde{N} : \pi_{M_1}(\widetilde{\theta}^{i_k}\varphi_{\mathrm{tr}}(\langle \xi_N\rangle_c)) \in \mathscr{A}^*_{1,r}\big\} \ge \frac{N}{M_1}(1-2\varepsilon)W_Q(\mathcal{U}_r), \quad r=1,\dots,R, \tag{2.119}$$

where $\varepsilon > 0$ can be made suitably small by tuning the choices of $\delta, \varepsilon_1, \varepsilon_2, \varepsilon_3$ above. (Re-read the constructions from [BGdH10], Sections 3.5.4 and 3.5.5, note that $\zeta \in \mathscr{A}^*_{3,r} \subset \langle \widetilde{E}_{\mathrm{tr}}\rangle_c^{M_3}$ contains at least $\approx M_3/M_2$ disjoint occurrences of subsentences from $\mathscr{A}^*_{2,r}$, and note that in turn each $\xi \in \mathscr{A}^*_{2,r} \subset \langle \widetilde{E}_{\mathrm{tr}}\rangle_c^{M_2}$ has the property $\mathrm{freq}_{\mathscr{A}^*_{1,r}}(\xi) \approx \varphi_{\mathrm{tr}}(\langle Q_r\rangle_c)(\mathscr{A}^*_{1,r}) \approx 1$, and therefore contains $\approx M_2/M_1$ disjoint occurrences of subsentences from $\mathscr{A}^*_{1,r}$.)

2.8.3 Estimating the cost of finding suitable stretches in the medium with the correct 'multi-colour statistics'

The fact that, for $N \gg M_3$ and cut-points $j_1 < \cdots < j_N$, $R^N_{j_1,\dots,j_N}(X) \in \mathcal{O}$ with $\mathcal{O}$ from (2.117) (recall (2.78)) has combinatorial consequences parallel to [BGdH10, Secion 3.5.5]. Namely, there are "good" loops (which are in the overwhelming majority because they come from inside patterns from some $\mathscr{A}^*_{1,r}$) and various types of "filling loops" (which are used for "sticking" consecutive subpatterns together, etc.). Finally, we need to estimate the entropy coming from the good loops, which will yield a factor $\approx \exp\big(NH_{\tau|K}([\langle Q\rangle_c]_{\mathrm{tr}})\big)$, and the "price" for finding substrings from the $\mathscr{B}_r$ in the medium X, which will yield a factor $\approx \exp(-N\alpha m_{[\langle Q\rangle_c]_{\mathrm{tr}}} H(\Psi_{[\langle Q\rangle_c]_{\mathrm{tr}}} \mid \langle \nu\rangle_c^{\otimes\mathbb{N}}))$. Furthermore, we have to control that the additional entropy terms coming from the various filling loops do not contribute in the limit. For this, we estimate $\mathbb{P}(R_N \in \mathcal{O} \mid X)$ similarly to (2.84) (for the ergodic case), and extend the estimate (2.92) from Section 2.6 to an "R-colour version". More precisely, we formulate an argument that has the same relation to [BGdH10, Lemma 3.3] as Section 2.6 has to [BGdH10, Lemmas 3.2 and 2.1]. Note that the approach taken in Section 2.6 is a bit more flexible than the original approach from [BGdH10, Section 3.5.5.B], because we do not need to coarse-grain to a fixed length scale.

For $r = 1, \dots, R$ define $\mathscr{B}_r = \{\widehat{\kappa}^*_{\mathrm{tr}}(y)\colon\, y \in \mathscr{A}^*_{1,r}\}$ analogous to (2.69) and for $z \in \mathscr{B}_r$ define $f^*_{\mathrm{tr},r}(z)$ analogous to (2.74) with $\mathscr{A}^*$ there replaced by $\mathscr{A}^*_{1,r}$. Then, analogous to (2.84), we have

$$\mathbb{P}(R_N \in \mathcal{O} \mid X) \le S_{N,R}, \tag{2.120}$$

where, similar to (2.83), we now have (recall $\widetilde{N}$ from (2.118))

$$\begin{aligned} S_{N,R} := \sum_{\substack{(r_1,\dots,r_{\widetilde{N}})\\ \text{compatible}}} \; \sum_{\substack{(z_1,\dots,z_{\widetilde{N}})\\ z_i \in \mathscr{B}_{r_i},\, i=1,\dots,\widetilde{N}}} \; \sum_{\substack{(p_{i,k})\\ \text{compatible}\\ \text{with } (z_i)}} \\ \sum_{\substack{(s_1,\dots,s_{\widetilde{N}}) \in \mathbb{N}_0^{\widetilde{N}}\\ s_1+\dots+s_{\widetilde{N}} \le N - M\widetilde{N}}} \; \prod_{i=1}^{\widetilde{N}} \Bigg(& \rho^{*s_i}\big(p_{i,1} - p_{i-1,\ell_{i-1}} - |\widetilde{z}_{i-1,\ell_{i-1}}|\big) f^*_{\mathrm{tr},r_i}(z_i) \\ & \times \prod_{k=1}^{\ell_i - 1} \Big(\mathbb{1}\big(\theta^{p_{i,k}}(X) \text{ begins with } \widetilde{z}_{i,k}\big) \rho\big(p_{i,k+1} - p_{i,k} - |\widetilde{z}_{i,k}|\big)\Big) \\ & \times\, \mathbb{1}\big(\theta^{p_{i,\ell_i}}(X) \text{ begins with } \widetilde{z}_{i,\ell_i}\big)\Bigg). \end{aligned} \tag{2.121}$$

The summation in (2.121) is subject to the following combinatorial constraints:

(1) The first sum runs over all "colourings" $(r_1, \dots, r_{\widetilde{N}}) \in \{1, \dots, R\}^{\widetilde{N}}$ that are compatible with the approximate ergodic decomposition of Q implicit in (2.111), in the sense that $\#\{i \le \widetilde{N} : r_i = r\} \ge \widetilde{N} W_Q(\widetilde{\mathcal{U}}_r)(1 - 2\varepsilon)$ for $r = 1, \dots, R$ (cf. (2.119)). Furthermore, by the definitions (2.115) of $\mathscr{A}^*_{3,r}$ and (2.113) of $\mathscr{A}^*_{2,r}$, we may assume that the sequence $r_1, \dots, r_{\widetilde{N}}$ contains many long "essentially monochromatic" substretches of length $\lfloor M_2/M_1 \rfloor$ in which at least a fraction $1 - 2\varepsilon_2$ of the indices have the same value. In fact, at least $(1-2\varepsilon_2)\widetilde{N}/(\lfloor M_2/M_1 \rfloor)$ of the substretches of length $\lfloor M_2/M_1 \rfloor$ must have this property. Furthermore, (2.117) enforces that the fraction of essentially monochromatic stretches with "colour" r must be at least $W_Q(\widetilde{\mathcal{U}}_r)(1 - 5\varepsilon_2)$ for $r = 1, \dots, R$.

(2) In the second sum, $z_i = (\widetilde{z}_{i,1}, \dots, \widetilde{z}_{i,\ell_i}) \in \mathscr{B}_{r_i}$ is composed of ℓ_i "stretches". Some $\widetilde{z}_{i,k}$ may be equal to the empty word ϵ, in which case we put $\mathbb{1}\big(\theta^{p_{i,k}}(X) \text{ begins with } \widetilde{z}_{i,k}\big) \equiv 1$.

(3) In the third sum, $(p_{i,k})$ is compatible with a given choice of $(z_1, \dots, z_{\widetilde{N}})$ when $p_{i,k+1} \ge p_{i,k} + |\widetilde{z}_{i,k}| + \mathrm{tr}$ for $k < \ell_i, i \le \widetilde{N}$ and $p_{i+1,1} \ge p_{i,\ell_i} + |\widetilde{z}_{i,\ell_i}|$ for $i < \widetilde{N}$ (analogous to (2.82) in the ergodic case). Here, $p_{i,k}$ describes where in the medium X the k-th stretch $\widetilde{z}_{i,k}$ of the i-th subword z_i is found.

(4) In the fourth sum, s_i describes how many words the empirical process cuts out between the occurrence of z_{i-1} and z_i. The condition $R^N_{j_1,\dots,j_N}(X) \in \mathcal{O}$ enforces that $s_1 + \dots + s_{\widetilde{N}} \le N - M_1\widetilde{N} \le \varepsilon N$ (see (2.118)–(2.119)).

As in Section 2.6, let us assume for the moment that ρ satisfies the bound (2.85) (which is stronger than (1.1)). Then we can use (2.86) and estimate

$$S_{N,R} \le \mathrm{e}^{\varepsilon(N)N} \sum_{\substack{(r_1,\dots,r_{\widetilde{N}})\\ \text{compatible}}} \; \sum_{\substack{(z_1,\dots,z_{\widetilde{N}})\\ z_i \in \mathscr{B}_{r_i},\, i=1,\dots,\widetilde{N}}} \; \sum_{\substack{(p_{i,k})\\ \text{compatible}\\ \text{with } (z_i)}} \; \prod_{i=1}^{\widetilde{N}} \Bigg(\big(p_{i,1} - p_{i-1,\ell_{i-1}} - |\widetilde{z}_{i-1,\ell_{i-1}}| + 1\big)^{-\alpha} f^*_{\mathrm{tr},r_i}(z_i)$$

$$\times \prod_{k=1}^{\ell_i-1} \Big(\mathbb{1}\big(\theta^{p_{i,k}}(X) \text{ begins with } \widetilde{z}_{i,k}\big)\big(p_{i,k+1} - p_{i,k} - |\widetilde{z}_{i,k}| + 1\big)^{-\alpha}\Big)$$

$$\times\ \mathbb{1}\Big(\theta^{p_{i,\ell_i}}(X) \text{ begins with } \widetilde{z}_{i,\ell_i}\Big)\Bigg) \tag{2.122}$$

with the same combinatorial constraints as in (1)–(3) below (2.121). Pick $b \in (1/\alpha, 1)$. As in Section 2.6, we have

$$\begin{aligned}
&\mathbb{E}[S_{N,R}^b] \\
&\le \mathrm{e}^{\varepsilon(N)bN} \sum_{\substack{(r_1,\dots,r_{\widetilde{N}})\\ \text{compatible}}} \ \sum_{\substack{(z_1,\dots,z_{\widetilde{N}})\\ z_i \in \mathscr{B}_{r_i},\, i=1,\dots,\widetilde{N}}} \ \sum_{\substack{(p_{i,k})\\ \text{compatible}\\ \text{with } (z_i)}} \prod_{i=1}^{\widetilde{N}} \Bigg(\Big(p_{i,1} - p_{i-1,\ell_{i-1}} - |\widetilde{z}_{i-1,\ell_{i-1}}| + 1\Big)^{-b\alpha} \big(f^*_{\mathrm{tr},r_i}(z_i)\big)^b \\
&\qquad \times \prod_{k=1}^{\ell_i-1} \Big(\mathbb{P}\big(\theta^{p_{i,k}}(X) \text{ begins with } \widetilde{z}_{i,k}\big)\big(p_{i,k+1} - p_{i,k} - |\widetilde{z}_{i,k}| + 1\big)^{-b\alpha}\Big) \\
&\qquad \times \mathbb{P}\Big(\theta^{p_{i,\ell_i}}(X) \text{ begins with } \widetilde{z}_{i,\ell_i}\Big)\Bigg) \\
&\le \exp\Bigg(\widetilde{N} M_1 b \sum_{r=1}^{R} W_Q(\mathcal{U}_r)(1-\delta_2)\Big(H_{\tau|K}([\langle Q_r\rangle_c]_{\mathrm{tr}}) \\
&\qquad + (1-\varepsilon_{\mathrm{tr}})\mathbb{E}_{[\langle Q_r\rangle_c]_{\mathrm{tr}}}\big[\log \rho(\tau_1) \,\big|\, \tau_1 < \mathrm{tr}\big] + 4\varepsilon_1\Big)\Bigg) \\
&\quad \times \Big(\sum_{p=1}^{\infty} p^{-b\alpha}\Big)^{\widetilde{N}} \mathrm{e}^{\varepsilon(N)bN} \sum_{\substack{(r_1,\dots,r_{\widetilde{N}})\\ \text{compatible}}} \ \sum_{\substack{(z_1,\dots,z_{\widetilde{N}})\\ z_i \in \mathscr{B}_{r_i},\, i=1,\dots,\widetilde{N}}} \Big(\sum_{p=\mathrm{tr}}^{\infty} p^{-b\alpha}\Big)^{\sum_{i=1}^{\widetilde{N}} \#\{j \le \ell_i : \widetilde{z}_{i,j} = \epsilon\}} \\
&\qquad \times \prod_{i=1}^{\widetilde{N}} \prod_{r=1}^{\ell_i} \mathbb{P}(X \text{ begins with } \widetilde{z}_{i,r}) \\
&\le \mathrm{e}^{\varepsilon(N)bN}\big(\zeta(b\alpha)\big)^{\widetilde{N}} \times \#\{(r_1,\dots,r_{\widetilde{N}}) \text{ compatible}\} \\
&\quad \times \exp\Bigg(\widetilde{N} M_1 \sum_{r=1}^{R} W_Q(\mathcal{U}_r)(1-\delta_2)\Big(b H_{\tau|K}([\langle Q_r\rangle_c]_{\mathrm{tr}}) + (1-\varepsilon_{\mathrm{tr}}) b \mathbb{E}_{[\langle Q_r\rangle_c]_{\mathrm{tr}}}\big[\log \rho(\tau_1) \,\big|\, \tau_1 < \mathrm{tr}\big] \\
&\qquad - m_{[\langle Q_r\rangle_c]_{\mathrm{tr}}} H(\Psi_{[\langle Q_r\rangle_c]_{\mathrm{tr}}} \mid \langle \nu\rangle_c^{\otimes \mathbb{N}}) + 5\varepsilon_1\Big)\Bigg) \\
&\le \exp\Bigg(\widetilde{\varepsilon}(N) N + b \sum_{r=1}^{R} W_Q(\mathcal{U}_r) I^{\mathrm{fin}}([\langle Q_r\rangle_c]_{\mathrm{tr}})\Bigg),
\end{aligned} \tag{2.123}$$

where $\varepsilon(N), \widetilde{\varepsilon}(N) > 0$ refer to terms that depend on N but can be made arbitrarily small by tuning the parameters in the construction above and taking N large. In the second and the third inequalities, we use the fact that, for $r = 1, \dots, R$, $\mathscr{B}_r$ fulfills the analogues of (2.75) and (2.72) with the same given ε_1 as fixed in the discussion below (2.109).

Arguing as in (2.90) and (2.91) and taking $b \downarrow 1/\alpha$, we find

$$\limsup_{N\to\infty} \frac{1}{N} \log S_{N,R} \le \delta - \sum_{r=1}^{R} W_Q(\mathcal{U}_r) I^{\mathrm{fin}}([\langle Q_r\rangle_c]_{\mathrm{tr}}) \le 2\delta - I^{\mathrm{fin}}([Q]_{\mathrm{tr}}) \quad \text{a.s.} \tag{2.124}$$

Again, as at the end of Section 2.6, we can now replace the assumption (2.85) by (1.1) analogous to the argument in [BGdH10, Section 3.6]. This completes the proof of (2.10) for general $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$.

A Auxiliary facts and observations

This appendix collects a few elementary lemmas that have been used along the way.

Lemma A.1. [Decomposition of entropy] *For $Q \in \mathcal{P}^{\mathrm{erg,fin}}(\widetilde{E}^{\mathbb{N}})$,*

$$H(\mathscr{L}_Q(\tau)) = H_{\tau|K}(Q) + I_Q(\tau||K), \tag{A.1}$$

where

$$I_Q(\tau||K) = \lim_{N\to\infty} \frac{1}{N} h\big(\mathscr{L}_Q(K_N, \tau^{(1,\dots,N)}) \,\big|\, \mathscr{L}_Q(K_N) \otimes \mathscr{L}_Q(\tau^{(1,\dots,N)})\big) \geq 0 \tag{A.2}$$

is the specific mutual information between $K = (K_N)_{N\in\mathbb{N}} = (\kappa(Y^{(1)}, \dots, Y^{(N)}))_{N\in\mathbb{N}}$ and τ under Q. (Note that $H_{\tau|K}(Q) - H(\mathscr{L}_Q(\tau)) \leq 0$ can be viewed as the drop in specific entropy of word lengths under Q caused by conditioning on a typical concatenation.)

Proof. Assume first that $|E| < \infty$. Then

$$\begin{aligned}
&\frac{1}{N} h\big(\mathscr{L}_Q(K_N, \tau^{(1,\dots,N)}) \,\big|\, \mathscr{L}_Q(K_N) \otimes \mathscr{L}_Q(\tau^{(1,\dots,N)})\big) \\
&= \frac{1}{N} \sum_{y^{(1)},\dots,y^{(N)} \in \widetilde{E}} Q(Y^{(i)} = y^{(i)},\, i = 1, \dots, N) \\
&\qquad\qquad \times \log \frac{Q(Y^{(i)} = y^{(i)},\, i = 1, \dots, N)}{Q(K_N = \kappa(y^{(1)}, \dots, y^{(N)}))Q(|Y^{(i)}| = |y^{(i)}|,\, i = 1, \dots, N)} \\
&= \frac{1}{N} \sum_{y^{(1)},\dots,y^{(N)} \in \widetilde{E}} Q(Y^{(i)} = y^{(i)},\, i = 1, \dots, N) \\
&\qquad\qquad \times \log Q(Y^{(i)} = y^{(i)},\, i = 1, \dots, N \mid K_N = \kappa(y^{(1)}, \dots, y^{(N)})) \\
&\quad - \frac{1}{N} \sum_{y^{(1)},\dots,y^{(N)} \in \widetilde{E}} Q(Y^{(i)} = y^{(i)},\, i = 1, \dots, N) \log Q(|Y^{(i)}| = |y^{(i)}|,\, i = 1, \dots, N).
\end{aligned} \tag{A.3}$$

As $N \to \infty$, the term in the last line converges to $H(\mathscr{L}_Q(\tau))$, By [B08, Lemma 3], the term in the next to last line converges to $-H_{\tau|K}(Q)$.

For general E, consider a nested sequence of colour-coarse-grainings $\langle\cdot\rangle_c$ as in Section 2.7, apply the above to $\langle Q\rangle_c \in \mathcal{P}^{\mathrm{erg,fin}}(\langle\widetilde{E}\rangle_c^{\mathbb{N}})$, and take $c \to \infty$ afterwards. □

Lemma A.2. [Finite relative entropy implies finite log-moment] *Let $\rho, \rho' \in \mathcal{P}(\mathbb{N})$, and assume that $h(\rho' \mid \rho) < \infty$ and $\rho(k) \leq Ck^{-a}$ for some $a > 1$. Then there are constants c_1, c_2 ($c_1 = c_1(\rho), c_2 = c_2(\rho) < \infty$) such that*

$$\sum_{n\in\mathbb{N}} \rho'(n) \log n \leq c_1 + c_2 h(\rho' \mid \rho). \tag{A.4}$$

In particular, $\sum_{n\in\mathbb{N}} \rho'(n) \log n < \infty$.

Proof. This bound is well known (see [B08, Lemma 7] for an analogous result when ρ has exponential tails). For completeness we include a short argument. By assumption,

$$\infty > h(\rho' \mid \rho) = \sum_{n\in\mathbb{N}\colon\, \rho'(n)>\rho(n)} \rho(n)\frac{\rho'(n)}{\rho(n)}\log\frac{\rho'(n)}{\rho(n)} + \sum_{n\in\mathbb{N}\colon\, \rho'(n)\le\rho(n)} \rho(n)\frac{\rho'(n)}{\rho(n)}\log\frac{\rho'(n)}{\rho(n)}. \tag{A.5}$$

Since $x\mapsto x\log x$ is continuous on $[0,1]$, the second sum takes some value $\in[-1/e,0]$ and thus, for $a'\in(1,a)$,

$$\begin{aligned}\frac{1}{e}+h(\rho'\mid\rho) \ge \sum_{n\in\mathbb{N}\colon\, \rho'(n)>\rho(n)} \rho'(n)\log\frac{\rho'(n)}{\rho(n)} &\ge \sum_{n\in\mathbb{N}\colon\, \rho'(n)\ge Cn^{-a'}} \rho'(n)\log\frac{Cn^{-a'}}{Cn^{-a}}\\ &= (a-a')\sum_{n\in\mathbb{N}\colon \rho'(n)\ge Cn^{-a'}} \rho'(n)\log n.\end{aligned} \tag{A.6}$$

Combined with

$$\sum_{n\in\mathbb{N}\colon\, \rho'(n)<Cn^{-a'}} \rho'(n)\log n \le C\sum_{n\in\mathbb{N}}\frac{\log n}{n^{a'}} < \infty \tag{A.7}$$

this gives the claim. $\square$

Lemma A.3. [Finite log-moment implies finite entropy] *There are $c_1,c_2<\infty$ such that, for any $\rho\in\mathcal{P}(\mathbb{N})$,*

$$\sum_{n\in\mathbb{N}}\rho(n)\log\frac{1}{\rho(n)} \le c_1 + c_2\sum_{n\in\mathbb{N}}\rho(n)\log n. \tag{A.8}$$

Proof. Again, this bound is well known. Here is an argument for the sake of completeness:

$$\begin{aligned}\sum_{n\in\mathbb{N}}\rho(n)\log\frac{1}{\rho(n)} &= \sum_{n\in\mathbb{N}\colon\, \rho(n)\le n^{-3}}\rho(n)\log\frac{1}{\rho(n)} + \sum_{n\in\mathbb{N}\colon\, \rho(n)>n^{-3}}\rho(n)\log\frac{1}{\rho(n)}\\ &\le \sum_{n\in\mathbb{N}\colon\, \rho(n)\le n^{-3}}\frac{1}{n^2}\rho(n)^{1/3}\log\frac{1}{\rho(n)} + \sum_{n\in\mathbb{N}\colon\, \rho(n)>n^{-3}}\rho(n)\log(n^3)\\ &\le \sup_{x\in(0,1]} x^{1/3}\log(1/x)\sum_{n\in\mathbb{N}}\frac{1}{n^2} + 3\sum_{n\in\mathbb{N}}\rho(n)\log n.\end{aligned} \tag{A.9}$$

$\square$

Remark A.4. [Upper bound in terms of the annealed rate function] By Lemmas A.2–A.3, there are constants $c_1,c_2<\infty$ such that

$$\mathbb{E}_Q[\log(\tau_1)]\vee h(\mathscr{L}_Q(\tau_1)) \le c_1 + c_2 h(\pi_1 Q|q_{\rho,\nu}) \le c_1 + c_2 H(Q\mid Q_{\rho,\nu}), \quad Q\in\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}}). \tag{A.10}$$

In particular, $H(Q\mid Q_{\rho,\nu})<\infty$ implies that $\mathbb{E}_Q[\log(\tau_1)]<\infty$ and $h(\mathscr{L}_Q(\tau_1))<\infty$. $\triangle$

Observation A.5. [Conditioning on a larger σ-algebra reduces expected entropy] *Let X,Y,Z be (say, discrete) random variables on some probability space, and assume that $Z=f(Y)$ for some deterministic function f. Then*

$$\mathbb{E}\Big[h\big(\mathscr{L}(X\,|\,Y)\big)\Big] \le \mathbb{E}\Big[h\big(\mathscr{L}(X\,|\,Z)\big)\Big]. \tag{A.11}$$

Proof. This inequality is standard. Because $Z = f(Y)$, we can write $\mathscr{L}(X \,|\, Z) = \sum_y \mathscr{L}(X \,|\, Y = y)P(Y = y \,|\, Z)$ and combine this relation with the fact that entropy is concave. ☐

Observation A.6. [Bounding expected entropy differences via relative entropy] *In the setting of Observation A.5, assume that*

$$\mathbb{E}\Big[h\big(\mathscr{L}(X \,|\, Y) \,\big|\, \mu\big)\Big] \leq \delta \tag{A.12}$$

for some fixed probability measure μ on the state space of X. Then

$$0 \leq \mathbb{E}\Big[h\big(\mathscr{L}(X \,|\, Z)\big)\Big] - \mathbb{E}\Big[h\big(\mathscr{L}(X \,|\, Y)\big)\Big] \leq \delta. \tag{A.13}$$

Proof. Given Y, we have

$$\begin{aligned} h\big(\mathscr{L}(X \,|\, Y) \,\big|\, \mu\big) &= \sum_x P(X = x \,|\, Y) \log \frac{P(X = x \,|\, Y)}{\mu(x)} \\ &= \sum_x P(X = x \,|\, Y) \log P(X = x \,|\, Y) - \sum_x P(X = x \,|\, Y) \log \mu(x) \\ &= -h\big(\mathscr{L}(X \,|\, Y)\big) - \mathbb{E}\big[\log \mu(X) \,\big|\, Y\big], \end{aligned} \tag{A.14}$$

and analogously

$$h\big(\mathscr{L}(X \,|\, Z) \,\big|\, \mu\big) = -h\big(\mathscr{L}(X \,|\, Z)\big) - \mathbb{E}\big[\log \mu(X) \,\big|\, Z\big]. \tag{A.15}$$

Since $\mathbb{E}[\mathbb{E}[\log \mu(X) \,|\, Z]] = \mathbb{E}[\mathbb{E}[\log \mu(X) \,|\, Y]] = \mathbb{E}[\log \mu(X)]$, we find

$$\mathbb{E}\Big[h\big(\mathscr{L}(X \,|\, Z)\big)\Big] - \mathbb{E}\Big[h\big(\mathscr{L}(X \,|\, Y)\big)\Big] = \mathbb{E}\Big[h\big(\mathscr{L}(X \,|\, Y) \,\big|\, \mu\big)\Big] - \mathbb{E}\Big[h\big(\mathscr{L}(X \,|\, Z) \,\big|\, \mu\big)\Big] \leq \delta. \tag{A.16}$$

☐

The following multivariate analogue of Observation A.6 for specific entropies holds.

Observation A.7. [Bounding expected specific entropy differences via relative entropy] *Let S_X, S_Y, S_Z be finite sets, and let $X = (X_n)_{n \in \mathbb{N}}$, $Y = (Y_n)_{n \in \mathbb{N}}$, $Z = (Z_n)_{n \in \mathbb{N}}$ be shift-invariant stochastic processes defined on some joint probability space with values in $S_X^{\mathbb{N}}$, $S_Y^{\mathbb{N}}$, $S_Z^{\mathbb{N}}$, respectively. Assume that*

$$\limsup_{n \to \infty} \frac{1}{n} \mathbb{E}\Big[h\Big(\mathscr{L}\big((X_1, \ldots, X_n) \,\big|\, (Y_1, \ldots, Y_n)\big) \,\Big|\, \mu^{\otimes n}\Big)\Big] \leq \delta \tag{A.17}$$

for some probability measure μ on S_X, and assume that Z is a coding of Y with window-halfwidth 0, i.e., $Z_n = f(Y_n), n \in \mathbb{N}$, for some deterministic function $f\colon S_Y \to S_Z$. Then

$$\limsup_{n \to \infty} \frac{1}{n} \bigg(\mathbb{E}\Big[h\Big(\mathscr{L}\big((X_1, \ldots, X_n) \,\big|\, (Z_1, \ldots, Z_n)\big)\Big)\Big] - \mathbb{E}\Big[h\Big(\mathscr{L}\big((X_1, \ldots, X_n) \,\big|\, (Y_1, \ldots, Y_n)\big)\Big)\Big] \bigg) \leq \delta. \tag{A.18}$$

Proof. For fixed n apply Observation A.6 to $X = (X_1, \ldots, X_n)$, $Y = (Y_1, \ldots, Y_n)$, $Z = (Z_1, \ldots, Z_n)$. Afterwards divide by n and let $n \to \infty$. ☐

Remark A.8. [Letter content of long words is typical] Consider $Q \in \mathcal{P}^{\mathrm{erg}}(\widetilde{E}^{\mathbb{N}})$ with $m_Q = \infty$ and $H(Q \mid Q_{\rho,\nu}) < \infty$. Write $q := \pi_1 Q$ and note that $h(q \mid q_{\rho,\nu}) < \infty$. Parametrising $\rho_q(\ell) := q(E^\ell)$, $q_\ell(\cdot) := q(\cdot \cap E^\ell)/\rho_q(\ell)$ (i.e., q_ℓ is the law of a single word conditioned on its length being ℓ), we have

$$\begin{aligned} h(q \mid q_{\rho,\nu}) &= \sum_{\ell\in\mathbb{N}} \sum_{x_1,\dots,x_\ell \in E} \rho_q(\ell) q_\ell(x_1,\dots,x_\ell) \log \frac{\rho_q(\ell) q_\ell(x_1,\dots,x_\ell)}{\rho(\ell)\nu(x_1)\cdots\nu(x_\ell)} \\ &= h(\rho_q \mid \rho) + \sum_{\ell\in\mathbb{N}} \rho_q(\ell) h(q_\ell \mid \nu^{\otimes \ell}) < \infty. \end{aligned} \tag{A.19}$$

Since $\sum_{\ell\in\mathbb{N}} \ell \rho_q(\ell) = \infty$ by assumption, the bound in (A.19) enforces that $\frac{1}{\ell} h(q_\ell \mid \nu^{\otimes \ell})$ must be small for most large ℓ. △

Remark A.9. [$\widetilde{E}$ **is Polish**] Let E be a complete and separable metric space with metric $d_E(\cdot,\cdot)$, and $\widetilde{E} = \cup_{n\in\mathbb{N}} E^n$. The metric

$$d_{\widetilde{E}}\big((x_i)_{i=1}^m, (y_i)_{i=1}^n\big) := |m-n| + \sum_{i=1}^{m\wedge n} d_E(x_i, y_j) \tag{A.20}$$

turns $\widetilde{E}$ into a complete and separable metric space. △

B Variations on Abramov's formula: Proofs of Lemmas 2.6–2.7 and (2.47)

B.1 General observations

Let S be a finite or countable set, $X = (X^{(i)})_{i\in\mathbb{N}}$ a stationary ergodic process with state space S, defined on some probability space $(\Omega, \mathscr{F}, Q)$. We assume that the specific entropy of the process X, written $H(\mathscr{L}_Q(X)) < \infty$, is finite. Let $\emptyset \neq S_0 \subset S$ with $Q(X^{(1)} \in S_0) > 0$.

(I) Erasing information on S_0. Let $* \notin S$ be an additional symbol, put $S_* := (S \setminus S_0) \cup \{*\}$, and define $\varphi\colon S \to S_*$ by

$$\varphi\colon S \ni x \mapsto \begin{cases} x, & x \notin S_0, \\ *, & x \in S_0. \end{cases} \tag{B.1}$$

Extend φ in the obvious way to a function $\varphi\colon S^N \to S_*^N$ for $N \in \mathbb{N}$, and to $\varphi\colon S^{\mathbb{N}} \to S_*^{\mathbb{N}}$ by coordinate-wise application. Put $Z = (Z^{(i)})_{i\in\mathbb{N}} := \varphi(X) = (\varphi(X^{(i)}))_{i\in\mathbb{N}}$. Note that then $H(\mathscr{L}_Q(Z)) \le H(\mathscr{L}_Q(X)) < \infty$.

We have

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}_Q\Big[h\big(\mathscr{L}_Q(X^{(1,\dots,N)} \,|\, Z^{(1,\dots,N)})\big)\Big] = H\big(\mathscr{L}_Q(X)\big) - H\big(\mathscr{L}_Q(Z)\big), \tag{B.2}$$

where, in accordance with the notation conventions introduced above Lemma 2.9, the quantity $h(\mathscr{L}_Q(X^{(1,\dots,N)} \,|\, Z^{(1,\dots,N)}))$ denotes the random variable $f_N(Z^{(1,\dots,N)})$ with $f_N : S_*^N \to [0,\infty)$ defined by

$$f_N(z) = -\sum_{x\in S^N \,:\, \varphi(x)=z} Q(X^{(1,\dots,N)} = x \,|\, Z^{(1,\dots,N)} = z) \log Q(X^{(1,\dots,N)} = x \,|\, Z^{(1,\dots,N)} = z). \tag{B.3}$$

To verify (B.2), write

$$\begin{aligned}
&\mathbb{E}_Q\big[f_N(Z^{(1,\dots,N)})\big] \\
&= -\sum_{z\in S_*^N} Q(Z^{(1,\dots,N)}=z) \\
&\qquad \sum_{x\in S^N\,:\,\varphi(x)=z} Q(X^{(1,\dots,N)}=x\,|\,Z^{(1,\dots,N)}=z)\log Q(X^{(1,\dots,N)}=x\,|\,Z^{(1,\dots,N)}=z) \\
&= -\sum_{z\in S_*^N}\ \sum_{x\in S^N\,:\,\varphi(x)=z} Q(X^{(1,\dots,N)}=x)\Big(\log Q(X^{(1,\dots,N)}=x)-\log Q(Z^{(1,\dots,N)}=z)\Big) \\
&= -\sum_{x\in S^N} Q(X^{(1,\dots,N)}=x)\log Q(X^{(1,\dots,N)}=x)+\sum_{z\in S_*^N} Q(Z^{(1,\dots,N)}=z)\log Q(Z^{(1,\dots,N)}=z) \\
&= h\big(\mathscr{L}_Q(X^{(1,\dots,N)})\big)-h\big(\mathscr{L}_Q(Z^{(1,\dots,N)})\big),
\end{aligned} \tag{B.4}$$

divide by N and take $N\to\infty$. Using ergodicity of X and Z, we get the claim in (B.2).

(II) The induced process. Define the induced process $\widehat{X}$ corresponding to X with base set S_0 analogously to the construction (2.15) in Section 2.3.1, i.e., $\widehat{X}=(\widehat{X}^{(i)})_{i\in\mathbb{N}}$ with

$$\widehat{X}^{(i)}=\big(X^{(T_i)},X^{(T_i+1)},\dots,X^{(T_{i+1}-1)}\big),\qquad i\in\mathbb{N}, \tag{B.5}$$

and $T_i=\inf\big\{n\in\mathbb{N}:\sum_{j=1}^n \mathbb{1}_{S_0}(X^{(j)})\geq i\big\}$ the time of the i-th return of X to S_0. Note that $\widehat{X}$ has state space $\widehat{S}=\bigcup_{m\in\mathbb{N}_0} S_0\times(S\setminus S_0)^m$. Note that, by the return time lemma, we have $T_i<\infty$ Q-a.s. for all i. Note that $\widehat{X}$ is well-defined and is ergodic (see [S96, Theorem. I.2.19] or [CFS82, Chapter 1, §5,Theorem 1]).

Abramov's formula [A59] relates the specific entropy of X and the specific entropy of $\widehat{X}$ via

$$H\big(\mathscr{L}_Q(X)\big)=Q(X^{(1)}\in S_0)H\big(\mathscr{L}_Q(\widehat{X})\big). \tag{B.6}$$

See e.g. [B76, Ch. IV, Theorem 4.14] or [K98, Exercise 3.2.23] for a proof (and also [CFS82, Chapter 1, §5, Theorem 1]). For a quick idea why (B.6) is true, note that by the covering exponent theorem (e.g. [S96, Theoremm I.7.4]), for $N\gg 1$ any set $A\subset S^N$ that is "typical" for $X^{(1,\dots,N)}$ under Q (with $Q(X^{(1,\dots,N)}\in A)\geq 1/2$, say) must have at least $\approx\exp(NH(\mathscr{L}_Q(X))$ elements. By the Kac lemma, a typical path of X of length N contains $N(Q(X^{(1)}\in S_0)+o(1))$ excursions, so by suitably removing initial and final pieces from such paths we can convert A into a set $\widehat{A}\subset\widehat{S}^{\lceil N(Q(X^{(1)}\in S_0)-\epsilon)\rceil}$ (with $0<\epsilon\ll 1$) that is typical for $\widehat{X}^{(1,\dots,\lceil N(Q(X^{(1)}\in S_0)-\epsilon)\rceil)}$. Furthermore, on the exponential scale, $\widehat{A}$ has the "same" size as A. Therefore, using the covering exponent theorem again, we must have $\exp(NH(\mathscr{L}_Q(X)))\approx\exp(N(Q(X^{(1)}\in S_0)H(\mathscr{L}_Q(\widehat{X})))$. Taking log, dividing by N and letting $N\to\infty$ (and controlling the error terms suitably), we find (B.6).

Combining (I) and (II). We can extend the mapping $\varphi\colon S\to S_*$ to $\widehat{S}$ via

$$\varphi((x_1,x_2,\dots,x_{\ell+1}))=(*,x_2,\dots,x_{\ell+1})\in\widehat{S}_*:=\cup_{m\in\mathbb{N}_0}\{*\}\times(S\setminus S_0)^m, \tag{B.7}$$

and also to a mapping $\varphi\colon\ \widehat{S}^{\mathbb{N}}\to\widehat{S}_*^{\mathbb{N}}$. Analogous to (B.2), we have

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}_Q\Big[h\big(\mathscr{L}_Q(\widehat{X}^{(1,\dots,N)}\,|\,\varphi(\widehat{X}^{(1,\dots,N)}))\big)\Big]=H\big(\mathscr{L}_Q(\widehat{X})\big)-H\big(\mathscr{L}_Q(\varphi(\widehat{X}))\big). \tag{B.8}$$

Combining this with (B.6), applied to both $\widehat{X}$ and $\varphi(\widehat{X})$, we obtain

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}_Q\Big[h\big(\mathscr{L}_Q(\widehat{X}^{(1,\dots,N)})\,|\,\varphi(\widehat{X}^{(1,\dots,N)})\big)\Big]=\frac{1}{Q(X^{(1)}\in S_0)}\Big(H\big(\mathscr{L}_Q(X)\big)-H\big(\mathscr{L}_Q(\varphi(X))\big)\Big). \tag{B.9}$$

B.2 Proof of Lemma 2.6

We first prove (2.18). Fix $\mathrm{tr}'\geq\mathrm{tr}$, put $S=\cup_{\ell=1}^{\mathrm{tr}'}E^\ell$ (note that $|S|<\infty$ because $|E|<\infty$) and $S_0=\cup_{\ell=\mathrm{tr}}^{\mathrm{tr}'}E^\ell$, use $Q=[Q]_{\mathrm{tr}'}$ in the construction from Section B.1, and let the word sequence Y take the role of X there. Then $[Q]_{\mathrm{tr}'}(Y^{(1)}\in S_0)=\varepsilon_{\mathrm{tr}}$ and (2.18) follows from (B.6).

For the formula in the second line of (2.17), fix $\mathrm{tr}\in\mathbb{N}$ and take $\mathrm{tr}'\geq\mathrm{tr}$. Then

$$\begin{aligned}H([Q]_{\mathrm{tr}'}\mid[Q_{\rho,\nu}]_{\mathrm{tr}'})&=-H([Q]_{\mathrm{tr}'})-\mathbb{E}_{[Q]_{\mathrm{tr}'}}\Bigg[\sum_{i=1}^{|Y^{(1)}|}\log\nu(Y_i^{(1)})\Bigg]\\&=-\varepsilon_{\mathrm{tr}}H\big(\widehat{[Q]_{\mathrm{tr}'}}^{(\mathrm{tr})}\big)-\varepsilon_{\mathrm{tr}}\mathbb{E}_{[Q]_{\mathrm{tr}'}}\Bigg[\sum_{j=1}^{\mathrm{tr}}\log\nu\big(\widehat{Y}_{1,j}^{(1)}\big)\Bigg]=\varepsilon_{\mathrm{tr}}H\big(\widehat{[Q]_{\mathrm{tr}'}}^{(\mathrm{tr})}\,\big|\,\widehat{[Q_{\rho,\nu}]_{\mathrm{tr}'}}^{(\mathrm{tr})}\big),\end{aligned} \tag{B.10}$$

where, for the second equality, we use (2.18) for the first term and argue as in (2.48) for the second term. Let $\mathrm{tr}'\to\infty$ to obtain the second line of (2.17). Note that this is a proper equality because $I^{\mathrm{ann}}(Q)<\infty$.

Finally, we can let $\mathrm{tr}'\to\infty$ in (2.18) to obtain the first line of (2.17). Note that for $m_Q=\infty$ it may happen that $H(Q)=\infty$, in which case the first line of (2.17) simply reads $\infty=\infty$.

B.3 Proof of Lemma 2.7

Let $(Y^{(i)})_{i\in\mathbb{N}}\sim[Q]_{\mathrm{tr}}$ and construct $(\widehat{Y}^{(i)})_{i\in\mathbb{N}}$ as in Section 2.3.1. Note that, by definition, $\kappa(\widehat{Y}^{(i)})$ is a deterministic function of $\widehat{Y}^{(i)}$. Therefore the fact that $\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})$ is ergodic follows from ergodicity of $\widehat{Q}^{(\mathrm{tr})}$, which was proved in Lemma 2.5. Furthermore, by construction, for $i\in\mathbb{N}$ the random word $\kappa(\widehat{Y}^{(i)})$ contains at most $(\mathrm{tr}-1)\mathsf{len}(\widehat{Y}^{(i)})+1$ letters. Hence (2.16) from Lemma 2.5 shows that the mean word length is finite under $\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})$.

Next, recall that a not necessarily shift-invariant probability measure P on $E^{\mathbb{N}}$ is called asymptotically mean stationary if

$$\overline{P}(G)=\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}P\big(\theta^{-k}G\big)\quad\text{exists for any measurable } G\subset E^{\mathbb{N}}. \tag{B.11}$$

Note that $\overline{P}\in\mathcal{P}^{\mathrm{inv}}(E^{\mathbb{N}})$, and $\overline{P}$ is called the stationary mean of P (see e.g. [G11, Section 1.7]).

Since $\Psi_{[Q]_{\mathrm{tr}}}\in\mathcal{P}^{\mathrm{erg}}(E^{\mathbb{N}})$ by [B08, Remark 5], we have for measurable $G\subset E^{\mathbb{N}}$, writing X for the random letter sequence,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mathbb{1}_G\big(\theta^kX\big)=\Psi_{[Q]_{\mathrm{tr}}}(G)\qquad\Psi_{[Q]_{\mathrm{tr}}}\text{-a.e. and in }\mathcal{L}^1(\Psi_{[Q]_{\mathrm{tr}}}). \tag{B.12}$$

The arguments in [B08, Section 3, Eq. (23)] imply that $\kappa([Q]_{\mathrm{tr}}) \ll \Psi_{[Q]_{\mathrm{tr}}}$. Hence we also have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_G\big(\theta^k \kappa(Y)\big) = \Psi_{[Q]_{\mathrm{tr}}}(G) \qquad [Q]_{\mathrm{tr}}\text{-a.e.} \tag{B.13}$$

By dominated convergence, the convergence in (B.13) occurs also in $\mathcal{L}^1([Q]_{\mathrm{tr}})$. Note that, by construction, under $[Q]_{\mathrm{tr}}$ the concatenation $\kappa\big((\kappa(\widehat{Y}_i^{(i)}))_{i\in\mathbb{N}}\big)$ can be obtained from $\kappa(Y)$ by a random shift

$$\kappa\big((\kappa(\widehat{Y}_i^{(i)}))_{i\in\mathbb{N}}\big) = \theta^S \kappa(Y), \tag{B.14}$$

where $S := \sum_{j<T_1} |Y^{(j)}|$ is $[Q]_{\mathrm{tr}}$-a.s. finite. Together, (B.13) and (B.14) yield that also

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_G\Big(\kappa\big((\kappa(\widehat{Y}_i^{(i)}))_{i\in\mathbb{N}}\big)\Big) = \Psi_{[Q]_{\mathrm{tr}}}(G) \qquad [Q]_{\mathrm{tr}}\text{-a.e.} \tag{B.15}$$

The fact that in (B.15) we again have convergence also in $\mathcal{L}^1([Q]_{\mathrm{tr}})$ by the dominated convergence theorem means that we indeed have $\overline{\kappa(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}}))} = \Psi_{[Q]_{\mathrm{tr}}}$. The fact that $\overline{\kappa([Q]_{\mathrm{tr}})} = \Psi_{[Q]_{\mathrm{tr}}}$ is established in [CdH13, Lemma 5.1], or can alternatively be argued for along the same lines as above. (Note that we could alternatively follow the arguments from [CdH13, Proof of Lemma 5.1] for $\kappa\big(\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})\big)$.)

To get (2.20), we note that $[Q]_{\mathrm{tr}}$-a.s.

$$m_{\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} |k(\widehat{Y}^{(i)})| = \lim_{N\to\infty} \frac{1}{N} \sum_{j=T_1}^{T_N-1} |Y^{(j)}| = \lim_{N\to\infty} \frac{T_N - T_1}{N} \frac{1}{T_N - T_1} \sum_{j=T_1}^{T_N-1} |Y^{(j)}|. \tag{B.16}$$

Since, under $[Q]_{\mathrm{tr}}$,

$$\lim_{N\to\infty} \frac{T_N - T_1}{N} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathsf{len}(\widehat{Y}^{(i)}) = \frac{1}{\varepsilon_{\mathrm{tr}}} \tag{B.17}$$

by Lemma 2.5 and

$$\lim_{M\to\infty} \frac{1}{M - T_1} \sum_{j=T_1}^{M} |Y^{(j)}| = \mathbb{E}_{[Q]_{\mathrm{tr}}}\big[|Y^{(1)}|\big] = m_{[Q]_{\mathrm{tr}}}, \tag{B.18}$$

we indeed get $m_{\widehat{\kappa}_{\mathrm{tr}}([Q]_{\mathrm{tr}})} = m_{[Q]_{\mathrm{tr}}}/\varepsilon_{\mathrm{tr}}$.

Now (2.21) follows from (2.20) and the fact that asymptotically mean stationary probability measures on $E^{\mathbb{N}}$ have the same specific entropies and specific relative entropies as their corresponding stationary means (see e.g. [G11, Theorem 3.1.1] and [G11, Lemma 7.5.1]).

B.4 Proof of (2.47)

Put $S = \cup_{\ell=1}^{\mathrm{tr}} E^\ell$, $S_0 = E^{\mathrm{tr}}$, use $Q = [Q]_{\mathrm{tr}'}$ in the construction from Section B.1. Then $[Q]_{\mathrm{tr}}(Y^{(1)} \in S_0) = \varepsilon_{\mathrm{tr}}$ and (2.47) follows from (B.9).

C Proofs of other results

C.1 Proof of (1.20) from Lemma 1.4

For a given $Q \in \mathcal{P}^{\mathrm{inv}}_{\mathrm{tr}}(\widetilde{E}^{\mathbb{N}})$, denote the term on the right-hand side of (1.20) by

$$\widehat{I}_{\mathrm{tr}}(Q) := \inf\left\{ I^{\mathrm{fin}}(Q') : \, Q' \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}}), [Q']_{\mathrm{tr}} = Q \right\}. \tag{C.1}$$

Obviously, $\widetilde{I}_{\mathrm{tr}}(Q) \leq \widehat{I}_{\mathrm{tr}}(Q)$ because $I^{\mathrm{fin}} = \widetilde{I}^{\mathrm{que}}$ on $\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}}) \subset \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$.

For the reverse inequality, pick $\delta > 0$. There exists $Q' \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{N}})$ with $I^{\mathrm{que}}(Q') \leq \widetilde{I}_{\mathrm{tr}}(Q) + \delta$ and $[Q']_{\mathrm{tr}} = Q$. By (2.12), $I^{\mathrm{que}}(Q') = \lim_{\mathrm{tr}' \to \infty} I^{\mathrm{fin}}([Q']_{\mathrm{tr}'})$. We can take $\mathrm{tr}' \geq \mathrm{tr}$ so large that $I^{\mathrm{fin}}([Q']_{\mathrm{tr}'}) \leq I^{\mathrm{que}}(Q') + \delta$. Since $[Q']_{\mathrm{tr}'} \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{N}})$ satisfies $[[Q']_{\mathrm{tr}'}]_{\mathrm{tr}} = Q$, we have $\widehat{I}_{\mathrm{tr}}(Q) \leq \widetilde{I}_{\mathrm{tr}}(Q) + 2\delta$. Now take $\delta \downarrow 0$.

C.2 Example for Remark 1.6: (1.25) is not equal to $\widetilde{I}_{\mathrm{tr}}(Q)$

Consider $E = \{0,1\}$, $\nu = (1-p)\delta_0 + p\delta_1$, $\mathrm{tr} = 1$ and $Q = (\delta_{(1)})^{\otimes\mathbb{N}}$. Then $\Psi_Q = (\delta_1)^{\otimes\mathbb{N}}$ and $H(Q \mid [q^{\otimes\mathbb{N}}_{\rho,\nu}]_{\mathrm{tr}}) + (\alpha-1) m_Q H(\Psi_Q \mid \nu^{\otimes\mathbb{N}}) = \alpha \log(1/p)$.

Define $H_0 := 0$, $H_i = \inf\{n > H_{i-1} : X_n = 1\}$, $i \in \mathbb{N}$. The event $A_N = \{H_i - H_{i-1} = \tau_i \text{ for } i = 1,\dots,N\}$ (with τ_i from (1.2)) implies $[R_N]_1 = Q$. Hence $\mathbb{P}(A_N \mid X) \leq \mathbb{P}([R_N]_1 \in \mathcal{O} \mid X)$ for any open neighbourhood $\mathcal{O} \ni Q$, and therefore $\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(A_N \mid X) \leq -\widetilde{I}_1(Q)$. Now write

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(A_N \mid X) = \lim_{N\to\infty} \frac{1}{N} \log \prod_{i=1}^{N} \rho(H_i - H_{i-1}) = \mathbb{E}\big[\log \rho(H_1)\big] = \sum_{k\in\mathbb{N}} p(1-p)^{k-1} \log \rho(k). \tag{C.2}$$

For fixed $p \in (0,1)$ we can construct examples of $\rho \in \mathcal{P}(\mathbb{N})$ satisfying (1.1) for which the expression on the right-hand side of (C.2) is strictly larger than $-\alpha \log(1/p)$.

References

[A59] L.M. Abramov, Entropy of an induced automorphism, Dokl. Acad. Sci. USSR 128, no. 4 (1959), 647–650.

[B65] P. Billingsley, *Ergodic Theory and Information*, Wiley, 1965.

[B08] M. Birkner, Conditional large deviations for a sequence of words, Stoch. Proc. Appl. 118, no. 5 (2008), 703–729.

[BGdH10] M. Birkner, A. Greven, F. den Hollander, Quenched large deviation principle for words in a letter sequence, Probab. Theory Relat. Fields 148, no. 3/4 (2010), 403–456.

[BGdH11] M. Birkner, A. Greven, F. den Hollander, Collision local time of transient random walks and intermediate phases in interacting stochastic systems, Electron. J. Probab. 16, no. 20 (2011), 552–586.

[BdH13] M. Birkner, F. den Hollander, A quenched large deviation principle in a continuous scenario, [arXiv:1312.2521], 2013.

[BdHO15] E. Bolthausen, F. den Hollander, A.A. Opoku, A copolymer near a selective interface: variational characterization of the free energy, Ann. Probab. 43, no. 2 (2015), 875-933.

[B57] L. Breiman, The individual ergodic theorem of information theory, Ann. Math. Statist. 28 (1957), 809–811.

[B76] J.R. Brown, *Ergodic Theory and Topological Dynamics*, Academic Press, 1976.

[CdH21] F. Caravenna, F. den Hollander, Phase transitions for spatially extended pinning, Probab. Theory Related Fields 181 (2021), no. 1-3, 329–375.

[CdH13] D. Cheliotis, F. den Hollander, Variational characterization of the critical curve for pinning of random polymers, Ann. Probab. 41, no. 3B (2013), 1767–1805.

[C61] K.L. Chung, A note on the ergodic theorem of information theory, Ann. Math. Statist. 32 (1961), 612–614.

[CFS82] I.P. Cornfeld, S.V. Fomin, Ya.G. Sinai, *Ergodic Theory*, Springer, 1982.

[DZ98] A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications* (2nd edition), Springer, 1998.

[DS89] J.-D. Deuschel, D.W. Stroock, *Large Deviations*, Academic Press, Boston, 1989.

[Ge88] H.-O. Georgii, *Gibbs Measures and Phase Transitions*, de Gruyter Studies in Mathematics 9, Walter de Gruyter, Berlin, 1988.

[G88] R.M. Gray, *Probability, Random Processes, and Ergodic Properties*, Springer, 1988.

[G11] R.M. Gray, *Entropy and Information Theory* (2nd edition), Springer, New York, 2011.

[dH00] F. den Hollander, *Large Deviations*, Fields Institute Monographs 14, American Mathematical Society, Providence, RI, 2000.

[dHO13] F. den Hollander, A.A. Opoku, Copolymer with pinning: variational characterization of the phase diagram, J. Stat. Phys. 152 (2013), no. 5, 846–893 .

[dHP14] F. den Hollander, J. Poisat, Large deviation principles for words drawn from correlated letter sequences, Electron. Commun. Probab. 19 (2014), no. 12, 16 pp.

[K98] G. Keller, *Equilibrium States in Ergodic Theory*, Cambridge University Press, 1998.

[MPW11] B. Marcus, K. Petersen, T. Weissman (editors), *Entropy of Hidden Markov Processes and Connections to Dynamical Systems*, London Mathematical Society Lecture Note Series 385, Cambridge, 2011.

[M17] J.C. Mourrat, Random PDEs: questions of regularity, Habilitation à diriger des recherches, Université de Lyon, 2017.

[N18] M. Nakashima, The free energy of the random walk pinning model, Stochastic Process. Appl. 128 (2018), no. 2, 373–403.

[S96] P.C. Shields, *The Ergodic Theory of Discrete Sample Paths*, Graduate Studies in Mathematics 13, American Mathematical Society, Providence RI, 1996.